PERCEPTRON LEARNING

David Kauchak
CS 158 – Fall 2025

1

Admin

Assignment 2 due Sunday at midnight

Slack (I *think* everyone is on the channel)

Additional mentor hours this week: Sat, 9-11am

Advice on testing

2

Bias

The “bias” of a model is how strong the model assumptions are.

low-bias classifiers make minimal assumptions about the data (*k*-NN and DT are generally considered low bias)

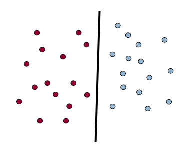
high-bias classifiers make strong assumptions about the data

3

Linear models

A strong high-bias assumption is *linear separability*:
- in 2 dimensions, can separate classes by a line
- in higher dimensions, need hyperplanes

A *linear model* is a model that assumes the data is linearly separable



4

Hyperplanes

A hyperplane is a line/plane in a high-dimensional space

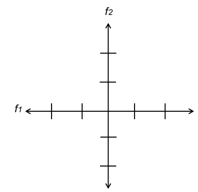
What defines a line?
What defines a hyperplane?

5

Defining a line

Any pair of values (*w1*,*w2*) defines a line through the origin:

$$0 = w_1 f_1 + w_2 f_2$$



6

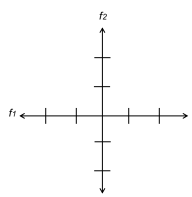
Defining a line

Any pair of values (*w1*,*w2*) defines a line through the origin:

$$0 = w_1 f_1 + w_2 f_2$$

$$0 = 1 f_1 + 2 f_2$$

-2	1
-1	0.5
0	0
1	-0.5
2	-1



7

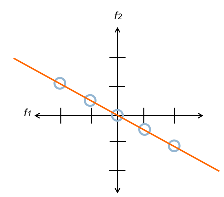
Defining a line

Any pair of values (*w1*,*w2*) defines a line through the origin:

$$0 = w_1 f_1 + w_2 f_2$$

$$0 = 1 f_1 + 2 f_2$$

-2	1
-1	0.5
0	0
1	-0.5
2	-1



8

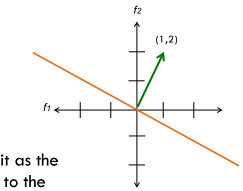
Defining a line

Any pair of values (w_1,w_2) defines a line through the origin:

$$0 = w_1 f_1 + w_2 f_2$$

$$0 = 1 f_1 + 2 f_2$$

w=(1,2)



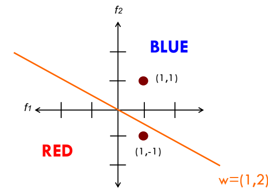
We can also view it as the line perpendicular to the *weight vector*

9

Classifying with a line

Mathematically, how can we classify points based on a line?

$$0 = 1 f_1 + 2 f_2$$



10

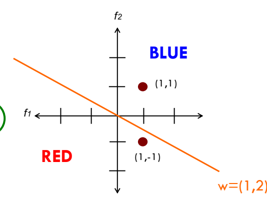
Classifying with a line

Mathematically, how can we classify points based on a line?

$$0 = 1 f_1 + 2 f_2$$

(1,1): $1*1+2*1 = 3$

(1,-1): $1*1+2*-1 = -1$



The sign indicates which side of the line

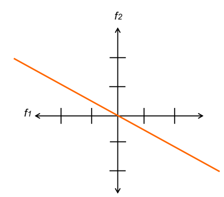
11

Defining a line

Any pair of values (w_1,w_2) defines a line through the origin:

$$0 = w_1 f_1 + w_2 f_2$$

$$0 = 1 f_1 + 2 f_2$$



How do we move the line off of the origin?

12

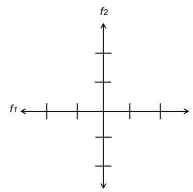
Defining a line

Any pair of values (w_1,w_2) defines a line through the origin:

$$a = w_1 f_1 + w_2 f_2$$

$$-1 = 1 f_1 + 2 f_2$$

f_1
-2
-1
0
1
2



13

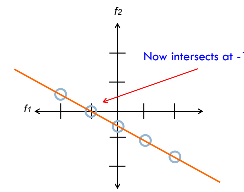
Defining a line

Any pair of values (w_1,w_2) defines a line through the origin:

$$a = w_1 f_1 + w_2 f_2$$

$$-1 = 1 f_1 + 2 f_2$$

f_1	f_2
-2	0.5
-1	0
0	-0.5
1	-1
2	-1.5



14

Linear models

A linear model in *n*-dimensional space (i.e. *n* features) is define by *n*+1 weights:

In two dimensions, a line:

$$0 = w_1 f_1 + w_2 f_2 + b \qquad \text{(where b = -a)}$$

In three dimensions, a plane:

$$0 = w_1 f_1 + w_2 f_2 + w_3 f_3 + b$$

In *n*-dimensions, a *hyperplane*

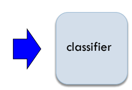
$$0 = b + \sum_{i=1}^{n} w_i f_i$$

15

Classifying with a linear model

We can classify with a linear model by checking the sign:

$f_1, f_2, \ldots, f_n$



$b + \sum_{i=1}^{n} w_i f_i > 0$ Positive example

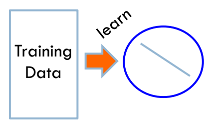
$b + \sum_{i=1}^{n} w_i f_i < 0$ Negative example

16

Learning a linear model

Geometrically, we know what a linear model represents

Given a linear model (i.e. a set of weights and b) we can classify examples

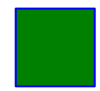


(data with labels)

How do we learn a linear model?

17

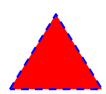
Positive or negative?



NEGATIVE

18

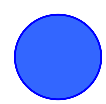
Positive or negative?



NEGATIVE

19

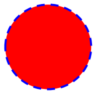
Positive or negative?



POSITIVE

20

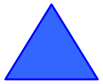
Positive or negative?



NEGATIVE

21

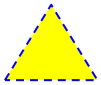
Positive or negative?



POSITIVE

22

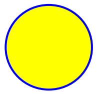
Positive or negative?



POSITIVE

23

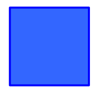
Positive or negative?



NEGATIVE

24

Positive or negative?



POSITIVE

25

A method to the madness

blue = positive

yellow triangles = positive

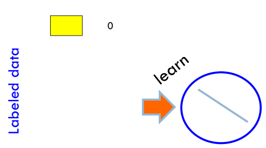
all others negative

How is this learning setup different than the learning we've seen so far?

When might this arise?

26

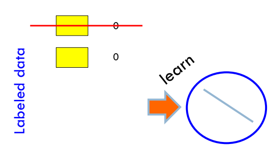
Online learning algorithm



Only get to see one example at a time!

27

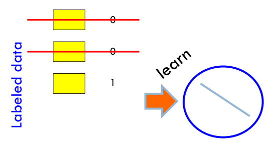
Online learning algorithm



Only get to see one example at a time!

28

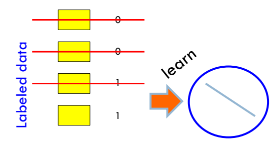
Online learning algorithm



Only get to see one example at a time!

29

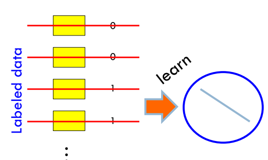
Online learning algorithm



Only get to see one example at a time!

30

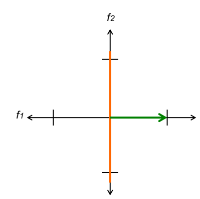
Online learning algorithm



Only get to see one example at a time!

31

Learning a linear classifier

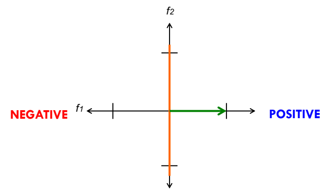


What does this model currently say?

w=(1,0)

32

Learning a linear classifier

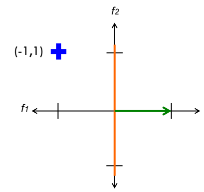


w=(1,0)

33

Learning a linear classifier

$0 = w_1 f_1 + w_2 f_2$



Is our current guess:
right or wrong?

w=(1,0)

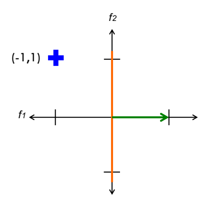
34

Learning a linear classifier

$0 = w_1 f_1 + w_2 f_2$

$1 * f_1 + 0 * f_2 =$

$1 * -1 + 0 * 1 = -1$



predicts negative, wrong

Geometrically, how should we update the model?

w=(1,0)

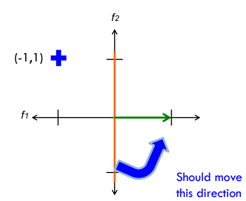
35

Learning a linear classifier

$0 = w_1 f_1 + w_2 f_2$

$1 * f_1 + 0 * f_2 =$

$1 * -1 + 0 * 1 = -1$



w=(1,0)

36

A closer look at why we got it wrong

w_1 w_2 (-1, 1, positive)

$$1 * f_1 + 0 * f_2 =$$

$$\underbrace{1 * -1 + 0 * 1}_{} = -1 \leftarrow$$

We'd like this value to be positive since it's a positive value

Which of the weights contributed to the mistake?

37

A closer look at why we got it wrong

w_1 w_2 (-1, 1, positive)

$$1 * f_1 + 0 * f_2 =$$

$$1 * -1 + 0 * 1 = -1 \leftarrow$$

We'd like this value to be positive since it's a positive value

contributed in the wrong direction

could have contributed (positive feature), but didn't

How should we change the weights?

38

A closer look at why we got it wrong

w_1 w_2 (-1, 1, positive)

$$1 * f_1 + 0 * f_2 =$$

$$1 * -1 + 0 * 1 = -1 \leftarrow$$

We'd like this value to be positive since it's a positive value

contributed in the wrong direction

could have contributed (positive feature), but didn't

decrease

1 -> 0

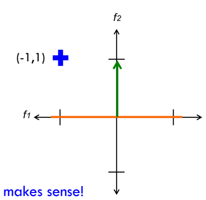
increase

0 -> 1

39

Learning a linear classifier

$$0 = w_1 f_1 + w_2 f_2$$



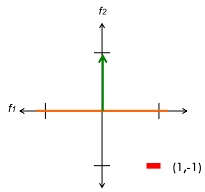
Geometrically, this also makes sense!

w=(0,1)

40

Learning a linear classifier

$0 = w_1 f_1 + w_2 f_2$



Is our current guess:
right or wrong?

w=(0,1)

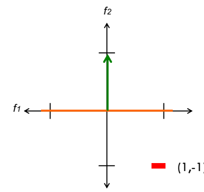
41

Learning a linear classifier

$0 = w_1 f_1 + w_2 f_2$

$0 * f_1 + 1 * f_2 =$

$0 * 1 + 1 * -1 = -1$



predicts negative, correct

How should we update the model?

w=(0,1)

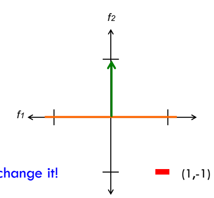
42

Learning a linear classifier

$0 = w_1 f_1 + w_2 f_2$

$0 * f_1 + 1 * f_2 =$

$0 * 1 + 1 * -1 = -1$



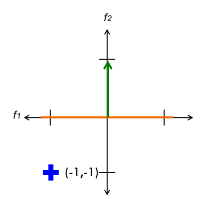
Already correct… don't change it!

w=(0,1)

43

Learning a linear classifier

$0 = w_1 f_1 + w_2 f_2$



Is our current guess:
right or wrong?

w=(0,1)

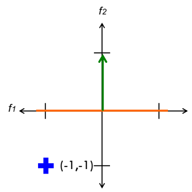
44

Learning a linear classifier

$0 = w_1 f_1 + w_2 f_2$

$0 * f_1 + 1 * f_2 =$

$0 * -1 + 1 * -1 = -1$



w=(0,1)

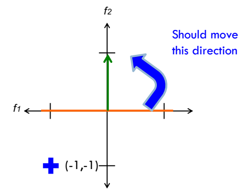
predicts negative, wrong

Geometrically, how should we update the model?

45

Learning a linear classifier

$0 = w_1 f_1 + w_2 f_2$



w=(0,1)

46

A closer look at why we got it wrong

w_1 w_2

(-1, -1, positive)

$0 * f_1 + 1 * f_2 =$

$0 * -1 + 1 * -1 = -1$

We'd like this value to be positive since it's a positive value

Which of the weights contributed to the mistake?

47

A closer look at why we got it wrong

w_1 w_2

(-1, -1, positive)

$0 * f_1 + 1 * f_2 =$

$0 * -1 + 1 * -1 = -1$

We'd like this value to be positive since it's a positive value

didn't contribute, but could have

contributed in the wrong direction

How should we change the weights?

48

A closer look at why we got it wrong

w_1 w_2

(-1, -1, positive)

$0 * f_1 + 1 * f_2 =$

$0 * -1 + 1 * -1 = -1$

We'd like this value to be positive since it's a positive value

didn't contribute, but could have

contributed in the wrong direction

decrease

0 -> -1

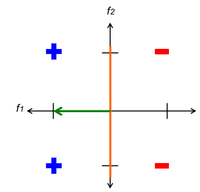
decrease

1 -> 0

49

Learning a linear classifier

f_1, f_2, label
-1,-1, positive
-1, 1, positive
1, 1, negative
1,-1, negative



w=(-1,0)

50

Perceptron learning algorithm

```
repeat until convergence (or for some # of iterations):
  for each training example (f1, f2, ..., fn, label):
    check if it's correct based on the current model

    if not correct, update all the weights:
      if label positive and feature positive:
        increase weight (increase weight = predict more positive)
      else if label positive and feature negative:
        decrease weight (decrease weight = predict more positive)
      else if label negative and feature positive:
        decrease weight (decrease weight = predict more negative)
      else if label negative and feature negative:
        increase weight (increase weight = predict more negative)
```

51

A trick…

	label * f_i
if label positive and feature positive:	1*1=1
increase weight (increase weight = predict more positive)	
else if label positive and feature negative:	1*-1=-1
decrease weight (decrease weight = predict more positive)	
else if label negative and feature positive:	-1*1=-1
decrease weight (decrease weight = predict more negative)	
else if label negative and negative weight:	-1*-1=1
increase weight (increase weight = predict more negative)	

52

A trick…

	label * f_i
if label positive and feature positive: increase weight (increase weight = predict more positive)	1*1=1
else if label positive and feature negative: decrease weight (decrease weight = predict more positive)	1*-1=-1
else if label negative and feature positive: decrease weight (decrease weight = predict more negative)	-1*1=-1
else if label negative and negative weight: increase weight (increase weight = predict more negative)	-1*-1=1

53

Perceptron learning algorithm

repeat until convergence (or for some # of iterations):
  for each training example (f_1, f_2, …, f_n, label):
    check if it's correct based on the current model

    if not correct, update all the weights:
      for each w_i:
        $w_i = w_i + f_i$*label
      $b = b$ + label

How do we check if it's correct?

54

Perceptron learning algorithm

repeat until convergence (or for some # of iterations):
  for each training example (f_1, f_2, …, f_n, label):
    $prediction = b + \sum_{i=1}^{n} w_i f_i$

    if *prediction* * *label* ≤ 0: // they don't agree
      for each w_i:
        $w_i = w_i + f_i$*label
      $b = b$ + label

55

Perceptron learning algorithm

repeat until convergence (or for some # of iterations):
  for each training example (f_1, f_2, …, f_n, label):
    $prediction = b + \sum_{i=1}^{n} w_i f_i$

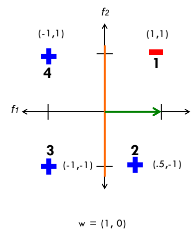
    if *prediction* * *label* ≤ 0: // they don't agree
      for each w_i:
        $w_i = w_i + f_i$*label
      $b = b$ + label

Would this work for non-binary features, i.e. real-valued?

56

Your turn ☺

f_1	f_2	label	w_1	w_2
			1	0
1	1	-1		
-1	1	1		
-1	-1	1		
.5	-1	1		
1	1	-1		
-1	1	1		
-1	-1	1		
.5	-1	1		
1	1	-1		
-1	1	1		
-1	-1	1		
.5	-1	1		

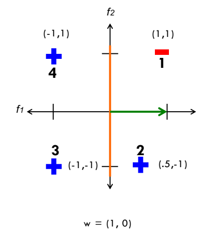


57

Your turn ☺

repeat until convergence (or for some # of iterations):
  for each training example (f_1, f_2, …, f_n, label):
    $prediction = \sum_{i=1}^{n} w_i f_i$
    if prediction * label ≤ 0: // they don't agree
      for each w_i:
        $w_i = w_i + f_i$*label

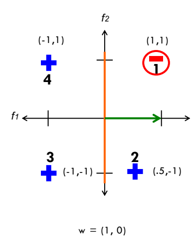
- Repeat until convergence
- Keep track of w_1, w_2 as they change
- Redraw the line after each step



58

Your turn ☺

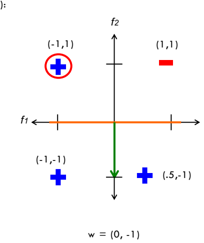
repeat until convergence (or for some # of iterations):
  for each training example (f_1, f_2, …, f_n, label):
    $prediction = \sum_{i=1}^{n} w_i f_i$
    if prediction * label ≤ 0: // they don't agree
      for each w_i:
        $w_i = w_i + f_i$*label



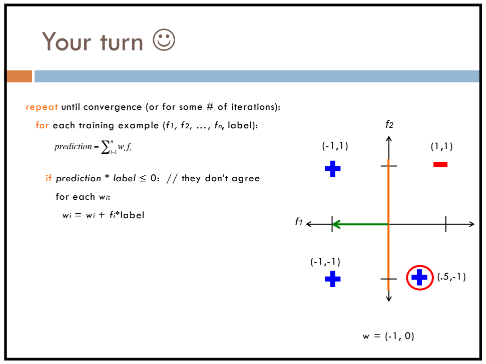
59

Your turn ☺

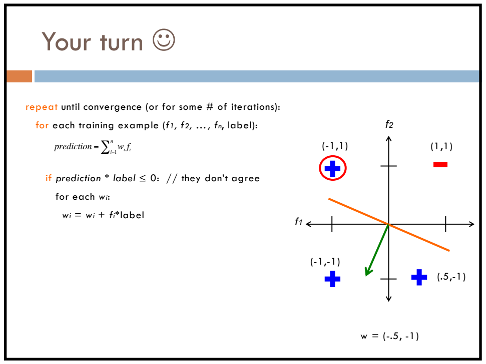
repeat until convergence (or for some # of iterations):
  for each training example (f_1, f_2, …, f_n, label):
    $prediction = \sum_{i=1}^{n} w_i f_i$
    if prediction * label ≤ 0: // they don't agree
      for each w_i:
        $w_i = w_i + f_i$*label



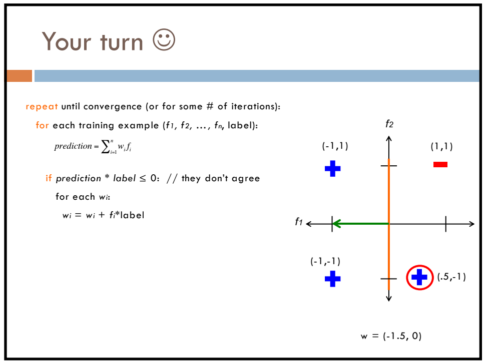
60


Your turn ☺
repeat until convergence (or for some # of iterations):
for each training example (f1, f2, …, fn, label):
prediction = $\sum_{i=1}^{n} w_i f_i$
if prediction * label ≤ 0: // they don't agree
for each wi:
wi = wi + fi*label
f2
f1
(-1,1)
(1,1)
(-1,-1)
(.5,-1)
w = (-1, 0)

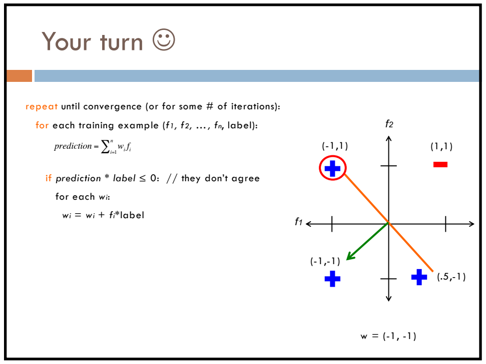
61


Your turn ☺
repeat until convergence (or for some # of iterations):
for each training example (f1, f2, …, fn, label):
prediction = $\sum_{i=1}^{n} w_i f_i$
if prediction * label ≤ 0: // they don't agree
for each wi:
wi = wi + fi*label
f2
f1
(-1,1)
(1,1)
(-1,-1)
(.5,-1)
w = (-.5, -1)

62


Your turn ☺
repeat until convergence (or for some # of iterations):
for each training example (f1, f2, …, fn, label):
prediction = $\sum_{i=1}^{n} w_i f_i$
if prediction * label ≤ 0: // they don't agree
for each wi:
wi = wi + fi*label
f2
f1
(-1,1)
(1,1)
(-1,-1)
(.5,-1)
w = (-1.5, 0)

63


Your turn ☺
repeat until convergence (or for some # of iterations):
for each training example (f1, f2, …, fn, label):
prediction = $\sum_{i=1}^{n} w_i f_i$
if prediction * label ≤ 0: // they don't agree
for each wi:
wi = wi + fi*label
f2
f1
(-1,1)
(1,1)
(-1,-1)
(.5,-1)
w = (-1, -1)

64

Your turn ☺

repeat until convergence (or for some # of iterations):

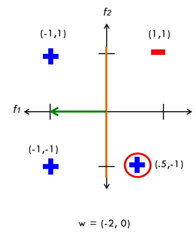
for each training example (f_1, f_2, …, f_n, label):

$prediction = \sum_{i=1}^{n} w_i f_i$

if $prediction * label \le 0$: // they don't agree

for each w_i:

$w_i = w_i + f_i$*label



65

Your turn ☺

repeat until convergence (or for some # of iterations):

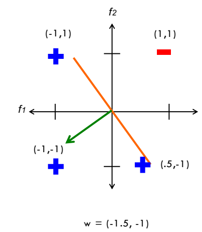
for each training example (f_1, f_2, …, f_n, label):

$prediction = \sum_{i=1}^{n} w_i f_i$

if $prediction * label \le 0$: // they don't agree

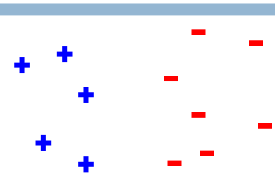
for each w_i:

$w_i = w_i + f_i$*label



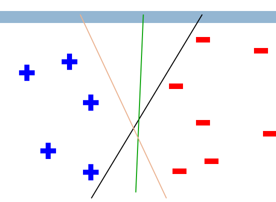
66

Which line will it find?



67

Which line will it find?



Only guaranteed to find ***some*** line that separates the data

68

Convergence

repeat until convergence (or for some # of iterations):

for each training example ($f_1, f_2, \ldots, f_n$, label):

$$prediction = b + \sum_{i=1}^{n} w_i f_i$$

if *prediction* * *label* ≤ 0: // they don't agree

for each w_i:

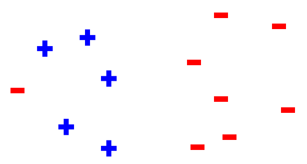
$w_i = w_i + f_i$*label

$b = b$ + label

Why do we also have the "some # iterations" check?

69

Handling non-separable data



If we ran the algorithm on this it would never converge!

70

Convergence

repeat until convergence (or for some # of iterations):

for each training example ($f_1, f_2, \ldots, f_n$, label):

$$prediction = b + \sum_{i=1}^{n} w_i f_i$$

if *prediction* * *label* ≤ 0: // they don't agree

for each w_i:

$w_i = w_i + f_i$*label

$b = b$ + label

Also helps avoid overfitting!
(This is harder to see in 2-D examples, though)

71

Ordering

repeat until convergence (or for some # of iterations):

for each training example ($f_1, f_2, \ldots, f_n$, label):

$$prediction = b + \sum_{i=1}^{n} w_i f_i$$

if *prediction* * *label* ≤ 0: // they don't agree

for each w_i:

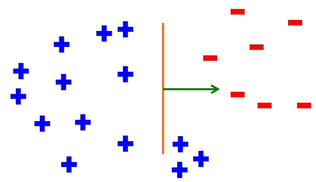
$w_i = w_i + f_i$*label

$b = b$ + label

What order should we traverse the examples?
Does it matter?

72

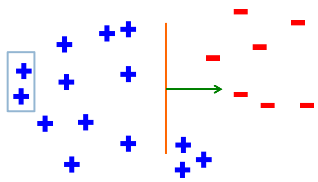
Order matters



What would be a good/bad order?

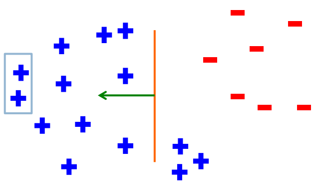
73

Order matters: a bad order



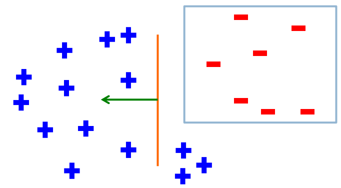
74

Order matters: a bad order



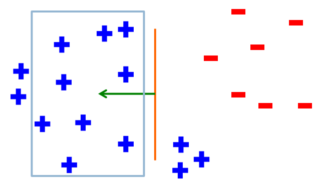
75

Order matters: a bad order



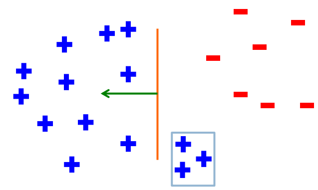
76

Order matters: a bad order



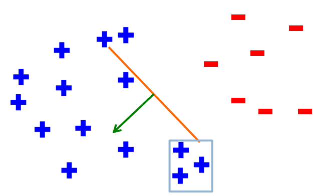
77

Order matters: a bad order



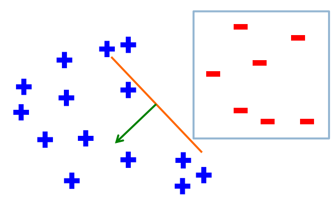
78

Order matters: a bad order



79

Order matters: a bad order



Solution?

80

Ordering

repeat until convergence (or for some # of iterations):

randomize order of training examples

for each training example (f_1, f_2, …, f_n, label):

$$prediction = b + \sum_{i=1}^{n} w_i f_i$$

if *prediction* * *label* ≤ 0: // they don't agree

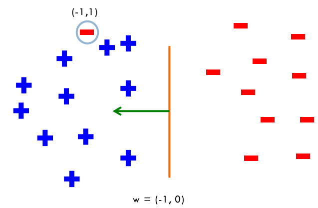
for each w_i:

$w_i = w_i + f_i$*label

b = b + label

81

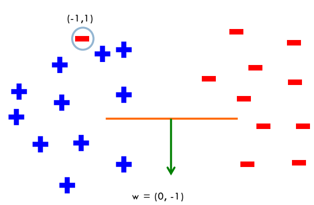
Improvements



What will happen when we examine this example?

82

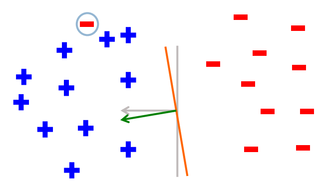
Improvements



Does this make sense? What if we had previously gone through ALL of the other examples correctly?

83

Improvements



Maybe just move it slightly in the direction of correction

84

Voted perceptron learning

Training

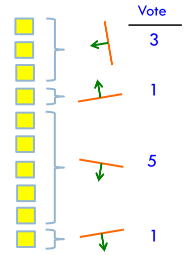
- every time a mistake is made on an example:
  - store the model weights (i.e. before changing for current example)
  - store the number of examples that set of weights got correct

Classify

- calculate the prediction from ALL saved weights
- multiply each prediction by the number it got correct (i.e., a weighted vote) and take the sum over all predictions
- said another way: pick whichever prediction has the most votes

85

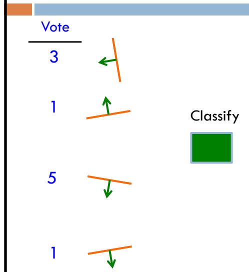
Voted perceptron learning



Training
every time a mistake is made on an example:
- store the weights
- store the number of examples that set of weights got correct

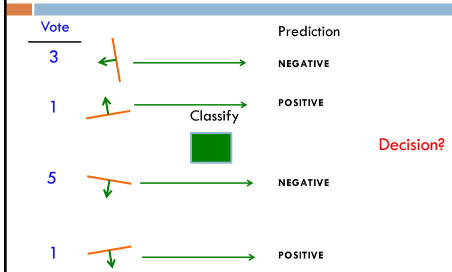
86

Voted perceptron learning



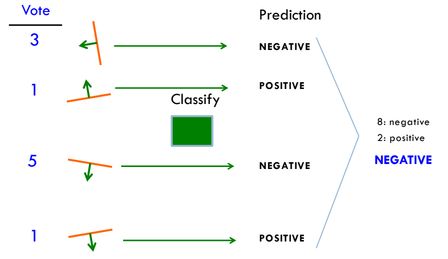
87

Voted perceptron learning



88

Voted perceptron learning



89

Voted perceptron learning

Works much better in practice

Avoids overfitting, though it can still happen

Avoids big changes in the result by examples examined at the end of training

90

Voted perceptron learning

Training
- every time a mistake is made on an example:
  - store the weights (i.e. before changing for current example)
  - store the number of examples that set of weights got correct

Classify
- calculate the prediction from ALL saved weights
- multiply each prediction by the number it got correct (i.e a weighted vote) and take the sum over all predictions
- said another way: pick whichever prediction has the most votes

Any issues/concerns?

91

Voted perceptron learning

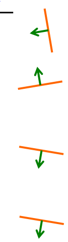
Training
- every time a mistake is made on an example:
  - store the weights (i.e. before changing for current example)
  - store the number of examples that set of weights got correct

Classify
- calculate the prediction from ALL saved weights
- multiply each prediction by the number it got correct (i.e a weighted vote) and take the sum over all predictions
- said another way: pick whichever prediction has the most votes

1. Can require a lot of storage
2. Classifying becomes very, very expensive

92

Average perceptron



Vote	
3	$w_1^1, w_2^1, \ldots, w_n^1, b^1$
1	$w_1^2, w_2^2, \ldots, w_n^2, b^2$
5	$w_1^3, w_2^3, \ldots, w_n^3, b^3$
1	$w_1^4, w_2^4, \ldots, w_n^4, b^4$

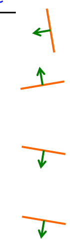
$$\overline{w_i} = \frac{3w_i^1 + 1w_i^2 + 5w_i^3 + 1w_i^4}{10}$$

The final weights are the *weighted average* of the previous weights

How does this help us?

93

Average perceptron



Vote	
3	$w_1^1, w_2^1, \ldots, w_n^1, b^1$
1	$w_1^2, w_2^2, \ldots, w_n^2, b^2$
5	$w_1^3, w_2^3, \ldots, w_n^3, b^3$
1	$w_1^4, w_2^4, \ldots, w_n^4, b^4$

$$\overline{w_i} = \frac{3w_i^1 + 1w_i^2 + 5w_i^3 + 1w_i^4}{10}$$

The final weights are the *weighted average* of the previous weights

Can just keep a running average!

94

Perceptron learning algorithm

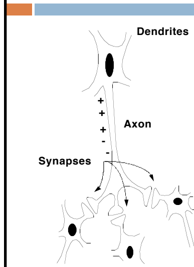
repeat until convergence (or for some # of iterations):
  for each training example ($f_1, f_2, \ldots, f_n$, label):
    $prediction = b + \sum_{i=1}^{n} w_i f_i$

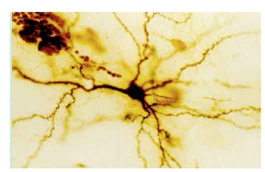
    if *prediction* * *label* ≤ 0: // they don't agree
      for each w_i:
        $w_i = w_i + f_i$*label
      b = b + label

Why is it called the "perceptron" learning algorithm if what it learns is a line? Why not "line learning" algorithm?

95

Our Nervous System

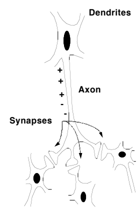




Neuron

96

Our nervous system: *the computer science view*

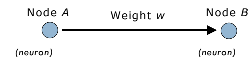


the human brain is a large collection of interconnected neurons

a NEURON is a brain cell

- collect, process, and disseminate electrical signals
- Neurons are connected via synapses
- They FIRE depending on the conditions of the neighboring neurons

97



w is the strength of signal sent between A and B.

If *A* fires and *w* is **positive**, then *A* **stimulates** *B*.

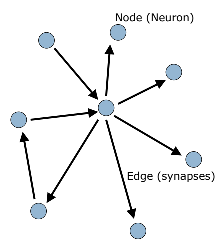
If *A fires* and *w* is **negative**, then *A* **inhibits** *B*.

If a node is stimulated enough, then it also fires.

How much stimulation is required is determined by its **threshold**.

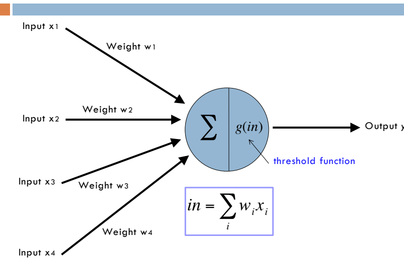
98

Neural Networks



99

A Single Neuron/Perceptron

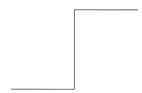


100

Possible threshold functions

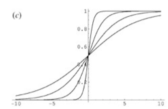
hard threshold:

if *in* (the sum of weights) >= *threshold* 1
else 0 otherwise



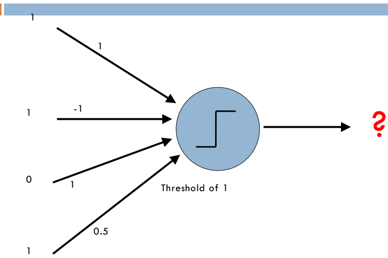
Sigmoid

$$g(x) = \frac{1}{1+e^{-ax}}$$



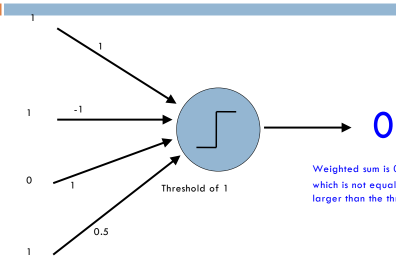
101

A Single Neuron/Perceptron



102

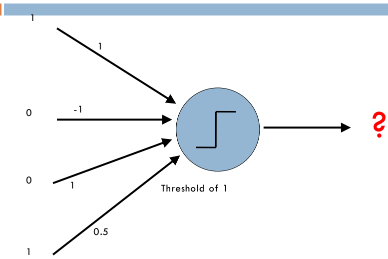
A Single Neuron/Perceptron



Weighted sum is 0.5, which is not equal or larger than the threshold

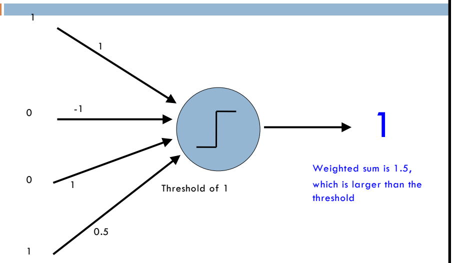
103

A Single Neuron/Perceptron



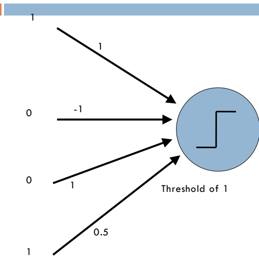
104

A Single Neuron/Perceptron



105

A Single Neuron/Perceptron



$$\sum_{i=1}^{n} w_i f_i > a$$

$$\text{b} + \sum_{i=1}^{n} w_i f_i > 0$$

where b = -a

106