Problem Set 6

Bookmark this page

Problem Sets due Jul 26, 2022 23:00 PKT Completed

Problems 1-5 correspond to "A simple linear classifier"

Problem 1

3/3 points (graded)

A linear classifier on $\mathbb{R}^2$ is specified by $w = (-1, 3)$ and $b = -6$.

a) At what point does the decision boundary intersect the x_1-axis? (Just give the x_1-intercept, a real number.)

-6

-6

b) At what point does the decision boundary intersect the x_2-axis? (Just give the x_2-intercept.)

2

2

c) What label, 1 or -1, is assigned to the point $(1, 1)$?

-1

-1

Submit

Problem 2

1/1 point (graded)

A particular line in $\mathbb{R}^2$ passes through the points $(0, 1)$ and $(2, 0)$ and is specified by equation $w \cdot x + b = 0$, where $b = -2$ and $w \in \mathbb{R}^2$. What is w?

- ○ $w = (0, 1)$
- ○ $w = (0, 2)$
- ◉ $w = (1, 2)$
- ○ $w = (2, -1)$

Submit

Problem 3

1/1 point (graded)

The Perceptron algorithm makes an update whenever it encounters a data point (x, y) that is "misclassified" by the current w, b. What does this mean, precisely? Choose the best option from this list.

- ( ) $y (w \cdot x + b) = 0$
- ( ) $y (w \cdot x + b) < 0$
- (•) $y (w \cdot x + b) \leq 0$
- ( ) $y (w \cdot x + b) > 0$

✓

Submit

Problem 4

1/1 point (graded)

A particular data set of n points is randomly permuted and then the Perceptron algorithm is run on it, repeatedly cycling through the points until convergence. It converges after k updates. Which of the following must be true? Select all that apply.

- [ ] $n \geq k$
- [x] If this process were repeated with a different random permutation, then the algorithm would again converge.
- [ ] If this process were repeated with a different random permutation, then the algorithm would again make k updates before convergence.
- [x] The data is linearly separable.

✓

Submit

Problem 5

1/1 point (graded)

The Perceptron algorithm is run on a data set, and converges after performing $p + q$ updates. Of these updates, p are on data points whose label is -1 and q are on data points whose label is $+1$. What is the final value of parameter b?

- (•) $q - p$
- ( ) $p + q$
- ( ) $p - q$
- ( ) q

✓

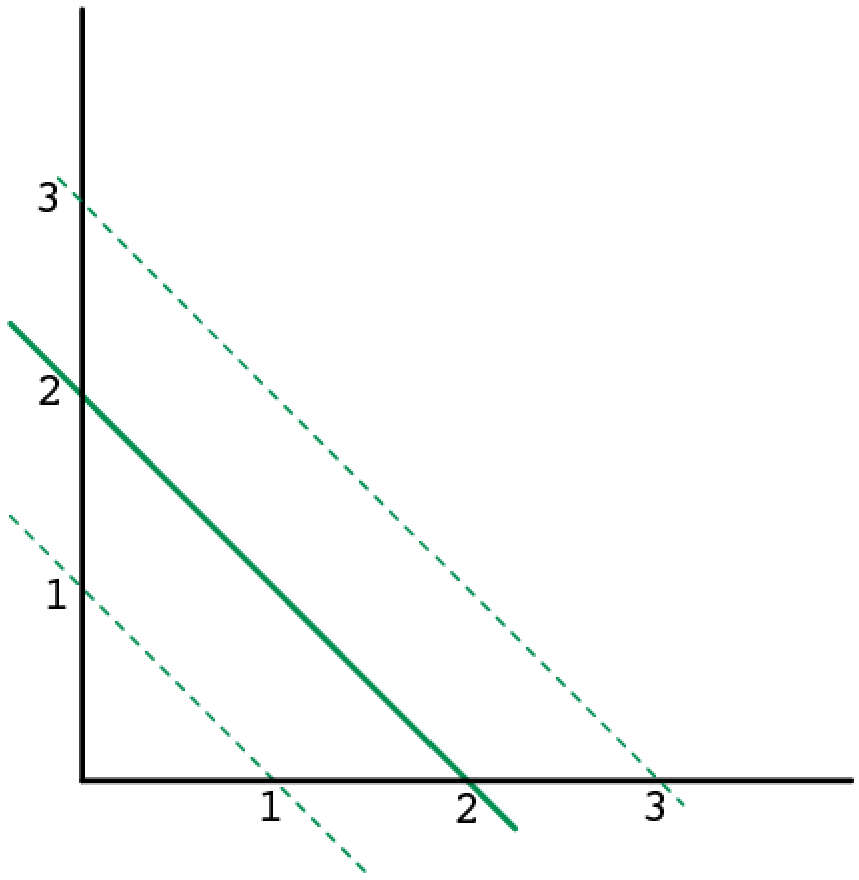
Submit

Problems 6-8 correspond to "Support vector machines I"

Problem 6

1/1 point (graded)

The figure below shows a two-dimensional linear separator $w \cdot x + b = 0$, along with the parallel lines $w \cdot x + b = -1$ and $w \cdot x + b = 1$.



What is the margin of this classifier?

- (selected) A number between 0.5 and 1.
- 1.
- A number between 1 and 2.
- 2.
- A number greater than 2.

✔

Submit

Problem 7

5/5 points (graded)

A support vector machine classifier is learned for a data set in $\mathbb{R}^2$. It is given by $w = (3, 4)$ and $b = -12$.

a) What is the x_1-intercept of the decision boundary?

4 ✔

4

b) What is the x_2-intercept of the decision boundary?

3

3

c) What is the margin of this classifier?

0.2

0.2

d) It turns out that the data set has two distinct support vectors of the form $(1, ?)$. What are they?

(give the missing x_2 coordinates for the support vectors with the smaller x_2 value first)

2

2

2.5

2.5

Submit

Problem 8

4/4 points (graded)

Consider the following small data set in $\mathbb{R}^2$:

Points $(1, 2)$, $(2, 1)$, $(2, 3)$, $(3, 2)$ have label -1.

Points $(4, 5)$, $(5, 4)$, $(5, 6)$, $(6, 5)$ have label $+1$.

Now, suppose (hard margin) SVM is run on this data.

a) What is the x_1-intercept of the decision boundary?

7

7

b) What is the x_2-intercept of the decision boundary?

7

7

c) What is w?

(1)

$w = \begin{pmatrix} \cdot \\ 2 \end{pmatrix}$

- ○ $w = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$
- ○ $w = \begin{pmatrix} 2 \\ 1/2 \end{pmatrix}$
- ◉ $w = \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix}$

✓

d) What is b?

-3.5 ✓

-3.5

Submit

Problems 9-12 correspond to "Support vector machines II"

Problem 9

4/4 points (graded)

Here is the optimization problem for the soft-margin SVM.

$$\min_{w\in\mathbb{R}^d,\, b\in\mathbb{R},\, \xi\in\mathbb{R}^n} \quad \|w\|^2 + C\sum_{i=1}^{n} \xi_i$$

$$\text{s.t.:} \quad y^{(i)}(w\cdot x^{(i)} + b) \geq 1 - \xi_i \quad \text{for all } i = 1, 2, \ldots, n$$

$$\xi \geq 0$$

a) How many slack variables are there? The answer should be a function of n and/or d.

n ✓

b) What setting of C will recapture the hard-margin SVM?

- ○ Very small C
- ◉ Very large C

( ) There is no value of C that will do this

Answer
Correct: Larger C imposes a heavier penalty on slack.

c) As C is increased, what happens to the margin of the linear classifier that is returned?

( ) The margin gets larger.

(•) The margin gets smaller.

( ) The margin is unchanged.

( ) The way in which the margin changes is unpredictable.

Answer
Correct:
As C grows, the optimization problem places more emphasis on classifying the training data correctly and less on having a big margin.

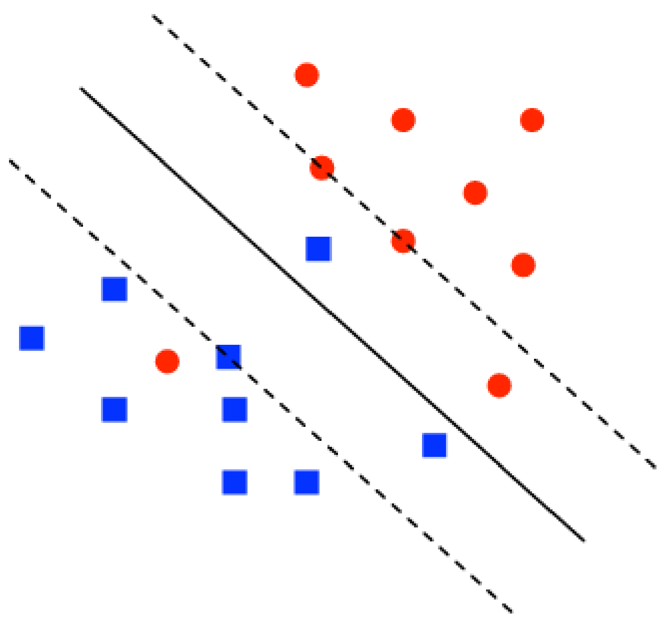
d) Suppose we have a data set that is linearly separable and we use it to train both a hard-margin SVM (w_H, b_H) and a soft-margin SVM (w_S, b_S). Which of the following statements is true? Select all that necessarily apply.

- [ ] Both linear classifiers have zero training error.
- [x] $\|w_H\| \geq \|w_S\|$
- [ ] $\|w_H\| \leq \|w_S\|$
- [x] The margin achieved by (w_H, b_H) is at most the margin achieved by (w_S, b_S).

Submit

Problem 10

4/4 points (graded)
The picture below shows the decision boundary obtained upon running soft-margin SVM on a small data set of blue squares and red circles.



a) How many support vectors are there?

7

7

b) What is the largest slack value on a red (circle-shaped) point, roughly?

2.3

2.3

c) What is the largest slack value on a blue (square-shaped) point, roughly?

1.2

1.2

d) Suppose the factor C in the soft-margin SVM optimization problem were increased. Would you expect the margin to **`increase`** or **`decrease`**?

decrease

Submit

Problem 11

1/1 point (graded)

Would it ever make sense to use the soft-margin SVM on a linearly separable data set? Select all that apply.

- [ ] No, unless you are unsure whether the data is linearly separable.
- [x] Yes, because it may lead to a larger margin and better generalization.
- [ ] No, because it might fail to perfectly separate the training set.

Submit

Problem 12

1/1 point (graded)

The soft-margin SVM involves a constant C that needs to be set. Which of the following is an appropriate way of setting it? Select all that apply.

- [ ] The output of the SVM is not very sensitive to the choice of C, so it doesn't really matter how C is set.
- [ ] Try various settings, and pick the one that yields the smallest training error.
- [x] Try various settings, and pick the one that yields the smallest cross-validation error.

- [x] Try various settings, and pick the one that yields the smallest cross-validation error.
- [ ] Try various settings, and pick the one that yields the largest margin.

Submit

Problems 13-16 correspond to "Duality"

Problem 13

1/1 point (graded)

The dual form of the Perceptron algorithm is run on a data set of four points $x \in \mathbb{R}^2$ with labels $y \in \{-1, 1\}$. The very first update takes place on the first data point, $x^{(1)} = (3, 2)$, which has label -1. What are the values of α and b right after this first update?

- ( ) $\alpha = (-1, 0, 0, 0)\ ,\ b = 1$
- ( ) $\alpha = (1, 0, 0, 0)\ ,\ b = 1$
- (•) $\alpha = (1, 0, 0, 0)\ ,\ b = -1$
- ( ) $\alpha = (0, 0, 0, 0)\ ,\ b = -1$

Submit

Problem 14

4/4 points (graded)

The dual form of the Perceptron algorithm is used to learn a binary classifier, based on n training points. It converges after k updates, and returns a vector α. For each of the following statements, indicate whether it is necessarily true or possibly false.

a) Each α_i is either 0 or 1.

possibly false

b) $\sum_i \alpha_i = k$.

necessarily true

c) α has at most k nonzero coordinates.

necessarily true

d) The training data must be linearly separable.

necessarily true

Submit

Problem 15

1/1 point (graded)

The dual form of the hard-margin SVM returns a vector α. Which data points $x^{(i)}$ are the support vectors in this solution?

- ( ) Those with $\alpha_i = 0$
- (•) Those with $\alpha_i > 0$
- ( ) Those with $\alpha_i \geq 0$
- ( ) The support vectors cannot be determined simply by looking at α

✓

Submit

Problem 16

2/2 points (graded)

Consider the primal and dual forms of the soft-margin SVM for binary classification. Suppose they are used on a training set of n points, where each point is d-dimensional.

a) How many real-valued variables are there in the primal optimization problem? (Don't use spaces in your expression.)

n+d+1 ✓

b) How many real-valued variables are there in the dual optimization problem?

n ✓

Submit

edX

About
Affiliates
edX for Business
Open edX
Careers
News

Legal

Terms of Service & Honor Code
Privacy Policy
Accessibility Policy
Trademark Policy
Sitemap

Connect

Blog
Contact Us
Help Center
Media Kit

GET IT ON Google play
Download on the App Store