# Problem 1

Consider a classification problem where we are given a training set of $n$ examples and labels $S_n = \{(x^{(i)}, y^{(i)}) : i = 1, \ldots, n\}$, where $x^{(i)} \in \mathbb{R}^2$ and $y^{(i)} \in \{1, -1\}$.

Assume a different data set for the two problems below.

## 1. (1)

2.0/2.5 points (graded)

Consider a classification problem where we are given a training set of $n$ examples and labels $S_n = \{(x^{(i)}, y^{(i)}) : i = 1, \ldots, n\}$, where $x^{(i)} \in \mathbb{R}^2$ and $y^{(i)} \in \{1, -1\}$.

Suppose for a moment that we are able to find a linear classifier with parameters $\theta'$ and $\theta_0'$ such that $y^{(i)} \left(\theta' \cdot x^{(i)} + \theta_0'\right) > 0$ for all $i = 1, \ldots, n$.

Let $\hat{\theta}$ and $\hat{\theta}_0$ be the parameters of the maximum margin linear classifier, if it exists, obtained by minimizing

$$\frac{1}{2}\|\theta\|^2 \qquad \text{subject to } \; y^{(i)}\left(\theta \cdot x^{(i)} + \theta_0\right) \geq 1 \; \text{ for all } \; i = 1, \ldots, n.$$

Determine if each of the following statements is True or False. (As usual, “True" means always true; “False" means not always true.)

1. The minimization problem defined by the equation immediately above has a solution if and only if the training examples $S_n$ are linearly separable.

- (•) True
- ( ) False

✓

2. The training examples $S_n$ are linearly separable under our assumptions.

- (•) True
- ( ) False

✓

3. $\left(\theta' \cdot x^{(i)} + \theta_0'\right) \leq \left(\hat{\theta} \cdot x^{(i)} + \hat{\theta}_0\right)$ for all $i = 1, \ldots, n$.

- ( ) True
- (•) False

✔

4. $\left(\theta' \cdot x^{(i)} + \theta'_0\right) \geq \left(\hat{\theta} \cdot x^{(i)} + \hat{\theta}_0\right)$ for all $i = 1, \ldots, n$.

- ( ) True
- (•) False

✔

5. $\|\theta'\| \geq \left\|\hat{\theta}\right\|$.

- (•) True
- ( ) False

✖

*Correction note (Sept 9):* The missing superscripts $(i)$ was added back to several $x$, in cases where the sentence says “for all $i = 1, \ldots, n$.

*Correction note (Sept 9):* The inequality sign in the optimization problem statement is fixed to be not strict. The earlier version was "subject to $y^{(i)}\left(\theta \cdot x^{(i)} + \theta_0\right) > 1$".

Submit You have used 3 of 3 attempts

## 1. (2)

2.0/4.0 points (graded)

Now we use kernel methods to classify a separate set of $n$ training examples (see figures below).

After trying out several methods, we generated 3 plots of $\left(\hat{\theta} \cdot \phi(x) + \hat{\theta}_0\right) = 0$ (solid), $\left(\hat{\theta} \cdot \phi(x) + \hat{\theta}_0\right) = 1$ (dashed), $\left(\hat{\theta} \cdot \phi(x) + \hat{\theta}_0\right) = -1$ (dashed), where $\hat{\theta}$ and $\hat{\theta}_0$ are the estimated ("primal") parameters.

Each plot was generated by optimizing the kernel version. In other words, we maximized

$$\sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i,j} \alpha_i \alpha_j y^{(i)} y^{(j)} K\left(x^{(i)}, x^{(j)}\right) \qquad \text{subject to } [\text{constraints on} \alpha_i]$$

with respect to $\alpha_i$ for $i = 1, \ldots, n$, where

$$\theta = \sum_{j=1}^{n} \alpha_j y^{(j)} \phi\left(x^{(j)}\right).$$

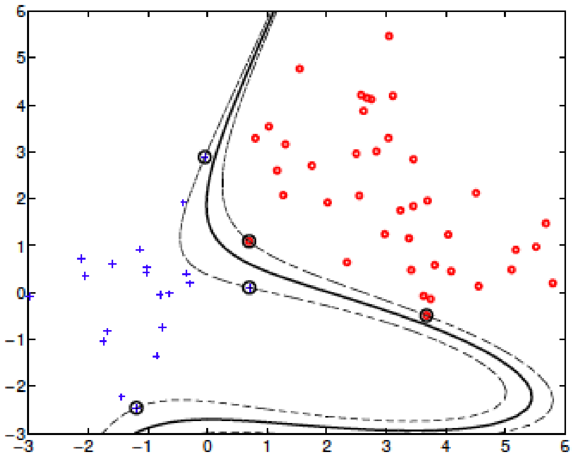
Each classifier was defined by a different choice of the kernel and the constraints.

Under each plot below, please identify a kernel-constraint pair (e.g., $(K_1, C_2)$) specifying the method that could have generated the plot.

**Note:** Each kernel could be associated to **at most 1 plot**.

*Correction Note (Sept 3):* In an earlier version, the problem contained an error, the plots $\left(\hat{\theta} \cdot \phi(x) + \hat{\theta}_0\right) = 0$ (solid), etc were written as $\left(\hat{\theta} \cdot x + \hat{\theta}_0\right) = 0$ etc.

*Correction Note (Sept 3):* In an earlier version, the relation $\theta = \sum_{j=1}^{n} \alpha_j y^{(j)} \phi\left(x^{(j)}\right)$ was assumed and not explicitly stated.

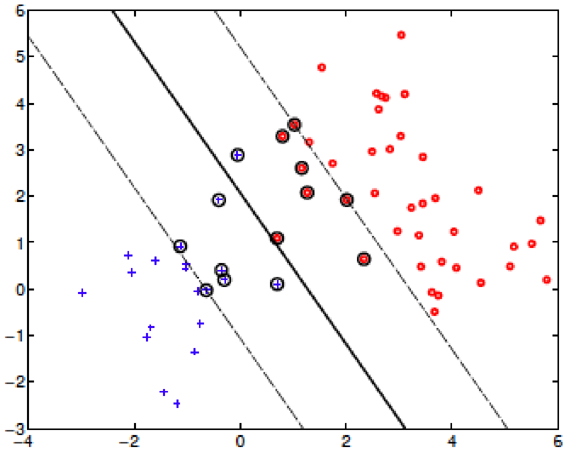



Kernel: Constraint:

(Select 1 per column.)

- $K_1(x, x') = (1 + x \cdot x'/2)$
- $K_2(x, x') = (1 + x \cdot x'/2)^2$
- $K_3(x, x') = (1 + x \cdot x'/2)^3$
- $K_g(x, x') = \exp\left(\|x \cdot x'\|^2/2\right)$

- $C_1 :\ 0 \leq \alpha_i \leq 0.1$ for all $i = 1, \ldots, n$
- $C_2 :\ \alpha_i \geq 0$ for all $i = 1, \ldots, n$



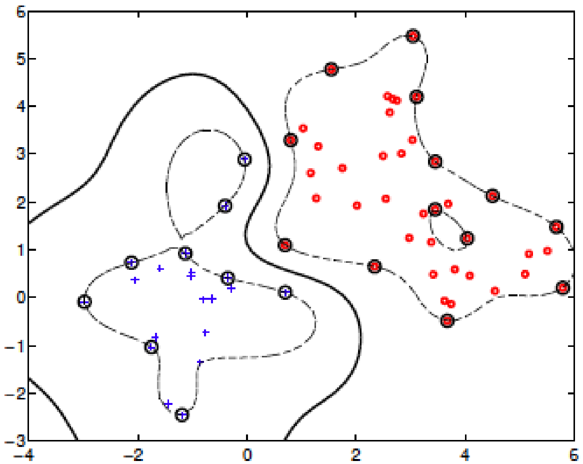
Kernel:

(Select 1 per column.)

- $K_1(x, x') = (1 + x \cdot x'/2)$
- $K_2(x, x') = (1 + x \cdot x'/2)^2$
- $K_3(x, x') = (1 + x \cdot x'/2)^3$
- $K_g(x, x') = \exp\left(\|x \cdot x'\|^2/2\right)$

Constraint:

- $C_1 : \; 0 \leq \alpha_i \leq 0.1$ for all $i = 1, \ldots, n$
- $C_2 : \; \alpha_i \geq 0$ for all $i = 1, \ldots, n$



Kernel:

(Select 1 per column.)

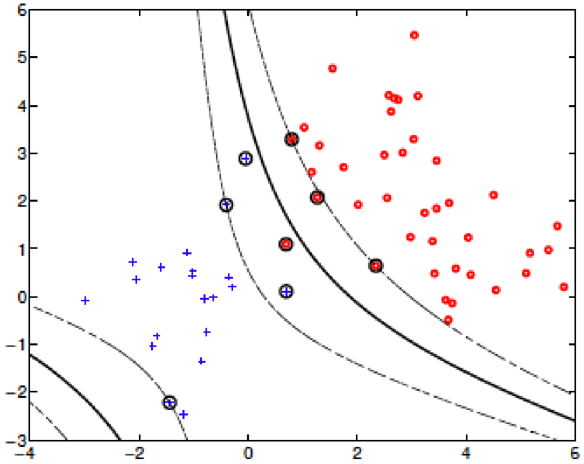
- $K_1(x, x') = (1 + x \cdot x'/2)$
- $K_2(x, x') = (1 + x \cdot x'/2)^2$
- $K_3(x, x') = (1 + x \cdot x'/2)^3$
- (selected) $K_g(x, x') = \exp\left(\|x \cdot x'\|^2/2\right)$

✔

Constraint:

- $C_1 : \ 0 \leq \alpha_i \leq 0.1$ for all $i = 1, \ldots, n$
- (selected) $C_2 : \ \alpha_i \geq 0$ for all $i = 1, \ldots, n$

✖




Kernel:

(Select 1 per column.)

Constraint:

- ( ) $K_1(x, x') = (1 + x \cdot x'/2)$
- (•) $K_2(x, x') = (1 + x \cdot x'/2)^2$
- ( ) $K_3(x, x') = (1 + x \cdot x'/2)^3$
- ( ) $K_g(x, x') = \exp\left(\|x \cdot x'\|^2/2\right)$

✔

- (•) $C_1 : \ 0 \leq \alpha_i \leq 0.1$ for all $i = 1, \ldots, n$
- ( ) $C_2 : \ \alpha_i \geq 0$ for all $i = 1, \ldots, n$

✔

Submit You have used 1 of 3 attempts

---

## Error and Bug Reports/Technical Issues

Hide Discussion

**Topic:** Final exam (1 week):Final Exam / Problem 1

Add a Post

< **All Posts**

### [STAFF]questions 1.1.4 and 1.1.5

question posted about 12 hours ago by rajivkbajpai

It appears that 1.4 and 1.5 are marked as wrong though they appear to be correct . In 1.3, the question was whether dot product of (x, theta_hat) is greater than (x,theta_dash) which is false and marked correctly by the grader but in 1.4 the opposite is marked as wrong i.e (x, theta_dash) is greater than (x,theta_hat). In 1.5, the norm of theta_hat should be the minimum as it is the params for max margin classifier, it should be less than than perceptron classifier but it is marked as wrong

Question 2 are also not correctly graded, it appears that graded answers are shifted down by one

This post is visible to everyone.

Add a Response

2 responses

**rajivkbajpai**
about 12 hours ago

+
...

In one of the discussion, it is stated that the theta_dash could be the params of soft margin classifier. It is not clear from the question whether the linear classifier defined could be soft margin classifier. If it is the case than norm of theta_hat could be more than that of soft margin classifier

Add a comment

**mrBB** (Community TA)
about 10 hours ago

+
...

I don't think there is a grading issue with 1. (1) (with 1. (2) there definitely is). The reason why I think $||\theta'||$ is not necessarily greater is than $||\hat{\theta}||$ is because $\theta'$ is just *a* separator for which $\theta' \cdot x + \theta'_0 \geq 1$ doesn't necessarily has to hold. So we can make $||\theta'||$ as large or small as we want by multiplying both $\theta'$ and $\theta'_0$ with an arbitrary positive constant, while keeping the same line/classifier.

And 1. (1) 3&4 don't hold for the same reason. Moreover, even in case we take $\theta' = c\hat{\theta}, \theta'_0 = c\hat{\theta}_0 \; c > 1$ then the statement doesn't hold for negatively classified points.

---

Thanks for the clarification.

posted about 10 hours ago by **rajivkbajpai**

---

@mrBB @rajivkbajpai I think we can argue like the following for Q1.(5)

Let's say the linear classifier is expressed as $\theta' \cdot x + \theta'_0 \geq \gamma > 0$ (note that such a $\gamma \in R^+$ can always be found). Now let's consider the following two (exhaustive) cases:

Case 1: $\gamma \geq 1 \Rightarrow \theta' \cdot x + \theta'_0 \geq 1 \Rightarrow ||\theta'|| \geq ||\hat{\theta}||$ (by max-margin)

Case 2: $0 < \gamma < 1$ in which case we can express the linear classifier as $\frac{\theta'}{\gamma} \cdot x + \frac{\theta'_0}{\gamma} \geq 1 \Rightarrow ||\frac{\theta'}{\gamma}|| \geq ||\hat{\theta}||$ (by max-margin) $\Rightarrow ||\theta'|| \geq \gamma.||\hat{\theta}||$. In this case, $||\theta'|| \geq ||\hat{\theta}||$ may not hold.

I thought like this initially and selected the correct answer, but later asked a question regarding the scaling of $||\theta'||$ (did not ask the right question) and got confused by the answer and ended up with selecting the wrong answer :-(.

posted 5 minutes ago by **sandipan_dey**

Add a comment

Preview

Submit

Showing all responses

## Add a response:

Preview

Submit