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)

2.5/2.5 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 θ' and θ_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 2 of 3 attempts

Answers are displayed within the problem

1. (2)

4.0/4.0 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 α_i for $i = 1, \ldots, n$, where

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

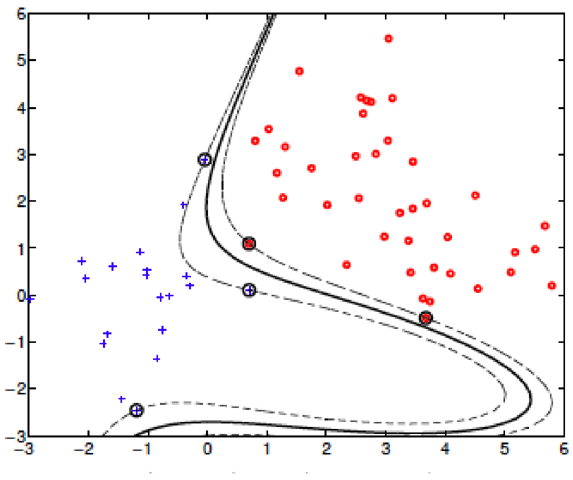
Each classifier was defined by a different choice of the kernel and the constraints.

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:

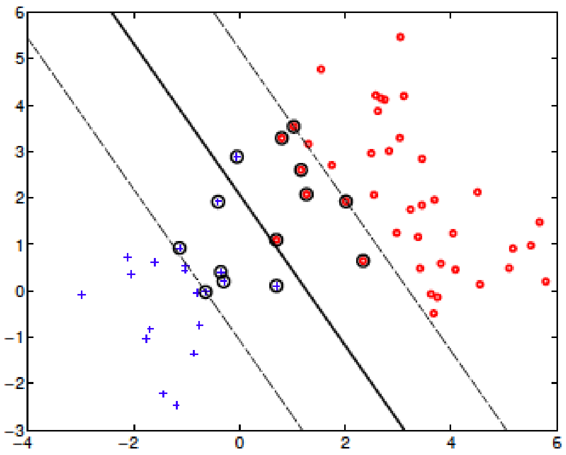
(Select 1 per column.)

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

✔

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

✔



Kernel:

Constraint:

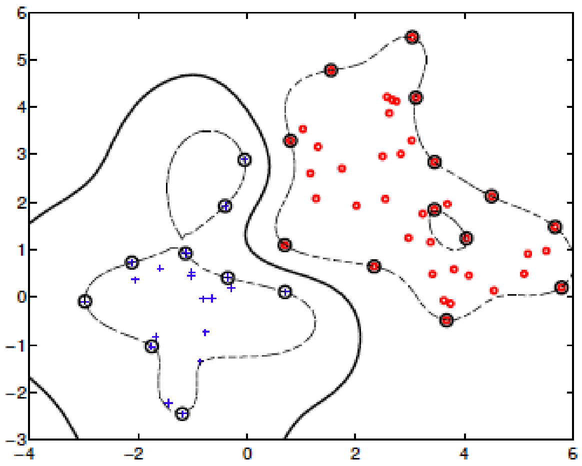
(Select 1 per column.)

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

✔

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

✔



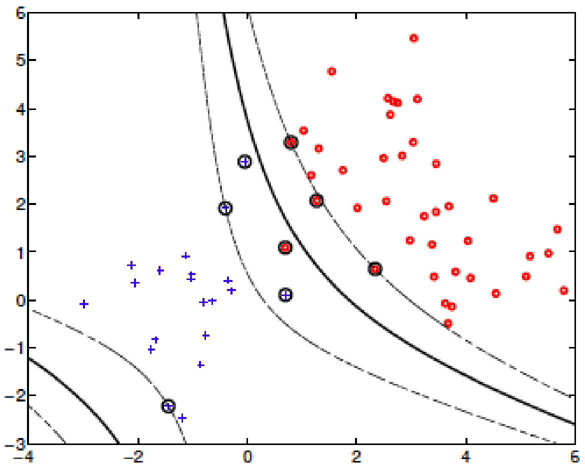
Kernel:

(Select 1 per column.)

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

Constraint:

- $C_1 : 0 \le \alpha_i \le 0.1$ for all $i = 1,\dots,n$
- (selected) $C_2 : \alpha_i \ge 0$ for all $i = 1,\dots,n$



Kernel:

(Select 1 per column.)

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

Constraint:

- (selected) $C_1 : 0 \le \alpha_i \le 0.1$ for all $i = 1,\dots,n$
- $C_2 : \alpha_i \ge 0$ for all $i = 1,\dots,n$

Submit You have used 2 of 3 attempts

Answers are displayed within the problem