CIS 419/519: Quiz 6

October 16, 2019

1. Stochastic gradient descent, when used with hinge loss, leads to which update rule?

   (a) Widrow's Adaline
   (b) Perceptron
   (c) Winnow
   (d) Adagrad

   Solution: (b)

2. Consider the following 4 data points:

   (i) $x_1 = [2, 2, -1]$
   (ii) $x_2 = [3, 3, -1]$
   (iii) $x_3 = [1, 0, -1]$
   (iv) $x_4 = [-2, -2, -2]$

   Assume we have some weight vector and bias:

   $w = [1, -2, 0], \theta = 0$

   Recall that the margin of a hyperplane is its distance to the closet point. The distance between a point x and the hyperplane defined by w and θ is: $\frac{w^T x + \theta}{||w||}$

   Which example x has the smallest margin?

   (a) x_1
   (b) x_2
   (c) x_3
   (d) x_4

   Solution: (c)

3. Given a kernel $k(x, y) = (x^T \cdot y + 3)^2$ where $x = [x_1, x_2]$ and $y = [y_1, y_2]$, which of the following is the correct representation of the kernel?

(a) $k(x, y) =< \phi(x), \phi(y) >$ where $\phi(x) = \begin{bmatrix} x_1^2 \\ x_2^2 \\ \sqrt{6}x_1 \\ \sqrt{6}x_2 \\ \sqrt{2}x_1x_2 \\ 3 \end{bmatrix}$, $\phi(y) = \begin{bmatrix} y_1^2 \\ y_2^2 \\ \sqrt{6}y_1 \\ \sqrt{6}y_2 \\ \sqrt{2}y_1y_2 \\ 3 \end{bmatrix}$

(b) $k(x, y) =< \phi(x), \phi(y) >$ where $\phi(x) = \begin{bmatrix} x_1^2 \\ x_2^2 \\ x_1x_2 \\ x_1 \\ x_2 \\ 3 \end{bmatrix}$, $\phi(y) = \begin{bmatrix} y_1^2 \\ y_2^2 \\ y_1y_2 \\ y_1 \\ y_2 \\ 3 \end{bmatrix}$

(c) $k(x, y) =< \phi(x), \phi(y) >$ where $\phi(x) = \begin{bmatrix} x_1^2 \\ x_2^2 \\ 3 \end{bmatrix}$, $\phi(y) = \begin{bmatrix} y_1^2 \\ y_2^2 \\ 3 \end{bmatrix}$

(d) None of the above. $k(x, y)$ is not a valid kernel

Solution: (a)

4. Given a kernel $k(x, y) = (x^T \cdot y + 3)^2$, what is the correct representation of the following kernel $k^{'}(x, y) = 3k(x, y)$?

(a) $k^{'}(x, y) =< \phi(x), \phi(y) >$ where $\phi(x) = \begin{bmatrix} \sqrt{3}x_1^2 \\ \sqrt{3}x_2^2 \\ 3\sqrt{2}x_1 \\ 3\sqrt{2}x_2 \\ \sqrt{6}x_1x_2 \\ 3^{\frac{3}{2}} \end{bmatrix}$, $\phi(y) = \begin{bmatrix} \sqrt{3}y_1^2 \\ \sqrt{3}y_2^2 \\ 3\sqrt{2}y_1 \\ 3\sqrt{2}y_2 \\ \sqrt{6}y_1y_2 \\ 3^{\frac{3}{2}} \end{bmatrix}$

(b) $k^{'}(x, y) =< \phi(x), \phi(y) >$ where $\phi(x) = \begin{bmatrix} 3x_1^2 \\ 3x_2^2 \\ 3\sqrt{6}x_1 \\ 3\sqrt{6}x_2 \\ 3\sqrt{2}x_1x_2 \\ 9 \end{bmatrix}$, $\phi(y) = \begin{bmatrix} 3y_1^2 \\ 3y_2^2 \\ 3\sqrt{6}y_1 \\ 3\sqrt{6}y_2 \\ 3\sqrt{2}y_1y_2 \\ 9 \end{bmatrix}$

(c) $k^{'}(x, y) =< \phi(x), \phi(y) >$ where $\phi(x) = \begin{bmatrix} 3x_1^2 \\ 3x_2^2 \\ 9 \end{bmatrix}$, $\phi(y) = \begin{bmatrix} 3y_1^2 \\ 3y_2^2 \\ 9 \end{bmatrix}$

(d) None of the above. $k^{'}(x, y)$ is not a valid kernel.

Solution: (a)

5. You are given a set of examples that are linearly inseparable over an original feature set X. Now we train two classifiers: (1) Classifier A is trained on this set of examples using a kernel equivalent to blowing up

the feature space to k dimensions (2) Classifier B is trained on this set of examples using a kernel equivalent to blowing up the feature space to n dimensions. If $k < n$, then Classifier B will always have a lower test error than Classifier A.

(a) True

(b) False

Solution: (b)