Homework 6

November 23, 2016

* For each question, give your answer. Justify your answer using a detailed explanation.

1 Understanding VC dimension

Generally, VC dimension is not related to the number of parameters used by a hypothesis, though we have seen that in linear and polynomial regression, more parameters imply higher VC dimension. Now, consider the hypothesis set

$$\mathcal{H} = \{f(x;\alpha) = \text{sign}(\sin(\alpha x)) | \alpha \in \mathbb{R}\}$$

for one dimensional classification.

(i) Show that $\mathcal{H}$ cannot shatter $m = 4$ points $x^{(1)} = 1, x^{(2)} = 2, x^{(3)} = 3, x^{(4)} = 4$.

(ii) Prove the VC dimension of $\mathcal{H}$ is ∞.

(Hints: for part (i), you need to find a set of labels $y^{(1)}, y^{(2)}, y^{(3)}, y^{(4)}$ such that for any α, $f(x;\alpha)$ cannot generate this set of labels, for example, you may want to use $+1, +1, -1, +1$; for part (ii), we have already shown it in class, you may want to use points $x^{(i)} = 10^{-i}, i = 1, \ldots, m$.)

2 Understanding SVM

Suppose we are given the following positively labeled data points in $\mathbb{R}^2$

$$\left\{ \begin{pmatrix} 3 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ -1 \end{pmatrix}, \begin{pmatrix} 6 \\ 1 \end{pmatrix}, \begin{pmatrix} 6 \\ -1 \end{pmatrix} \right\}.$$

and the following negatively labeled data points in $\mathbb{R}^2$

$$\left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \end{pmatrix} \right\}.$$

(i) Calculate the projection (in Euclidean distance) of a point $\bar{x}$ onto the hyperplane $\Omega = \{x \in \mathbb{R}^n \mid w^T x + b = 0\}$, i.e., find

$$\mathbb{P}_\Omega(\bar{x}) = \arg\min_x \{x \mid x \in \Omega\}.$$

What's the distance between $\bar{x}$ and $\mathbb{P}_\Omega(\bar{x})$?

(Hint: you may want to use KKT conditions.)

(ii) For the margin hyperplanes

$$\{x \in \mathbb{R}^n \mid w^T x + b = 1\}, \quad \text{and} \quad \{x \in \mathbb{R}^n \mid w^T x + b = -1\},$$

calculate the (Euclidian) distance between two hyperplanes.

(iii) Derive the primal problem, and use Matlab QP (quadratic programming) solver to find the solution. What is your optimal solution? What is your margin?

(iv) Derive the dual problem, and use Matlab QP solver to attack your dual problem. What is your dual optimal solution? What is your classifier? How would you classify point $(4,5)^T$?

(v) If you were required to do a leave-one-out cross validation, what would be the upper bound for your cross validation error?

Suppose we are given the following positively labeled data points in $\mathbb{R}^2$

$$\left\{ \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ -2 \end{pmatrix}, \begin{pmatrix} -2 \\ -2 \end{pmatrix}, \begin{pmatrix} -2 \\ 2 \end{pmatrix} \right\}.$$

and the following negatively labeled data points in $\mathbb{R}^2$

$$\left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \end{pmatrix} \right\}.$$

Use nonlinear mapping Φ from input space to some new feature space, where

$$\Phi\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{cases} \begin{pmatrix} 4 - x_2 + |x_1 - x_2| \\ 4 - x_1 + |x_1 - x_2| \end{pmatrix} & \text{if} \quad \sqrt{x_1^2 + x_2^2} > 2 \\ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} & \text{otherwise} \end{cases}$$

(i) Derive the primal problem for the given data set using this transformation, and use Matlab QP (quadratic programming) solver to find the solution. What is your optimal solution?

(ii) Write down the Kernel. Derive the dual problem for the given data set, and use Matlab QP solver to attack your dual problem. What is your dual optimal solution? What is your classifier? How would you classify point $(4, 5)^T$?