CSE512 Fall 2018 Machine Learning - Homework 4

Your Name: Caitao Zhan

Solar ID: 111634527

NetID email address: caitao.zhan@stonybrook.edu

Names of people whom you discussed the homework with:

1 Support Vector Machines

1.1 Linear case

Assume we haved learned the α's and b in the original input space

$$\mathbf{w}^T\mathbf{x} + b = \sum_{i=1}^{n} \alpha_i y^i \left\langle \mathbf{x}^i, \mathbf{x} \right\rangle + b \tag{1}$$

The prediction function of linear SVM:

$$f(\mathbf{x}; \mathbf{w}, b) = \begin{cases} 1 & iff\ (1) > 0 \\ -1 & otherwise \end{cases}$$

LOOCV error:

$$\frac{1}{n} \sum_{i=1}^{n} \delta(y^i, f(\mathbf{x}; \mathbf{w}^{-i}, b^{-i}))$$

where the superscript $-i$ denotes the parameters we found by removing the ith training example, and δ is an indicator function. Consider two cases:

1. Removing a support vector data point. The ith data point lies on the margin, and might be classified wrong. Because for such points, $\alpha_i > 0$, and might affect equation (1).

2. Removing a non-support vector data point. The ith data point lies outside the margin, and will be classified correctly for sure. Because for such points, $\alpha_i = 0$, and will not affect equation (1)

For case NO. 1, let's consider the worst case, that all m support vectors are classified wrong. This worst case leads to the upper bound of the LOOCV error $= \frac{m}{n}$

1.2 General case

The bound will still hold.

The definition of a kernel:

$$K(x, z) = \phi(x)^T \phi(z)$$

Then, everywhere we previously had $\langle \mathbf{x}, \mathbf{z} \rangle$ in our algorithm, we replace it with $K(x, z)$

Now assume we haved learned the (new) α's and b in the high dimensional feature space by using the kernel trick.

$$\mathbf{w}^T \phi(\mathbf{x}) + b = \sum_{i=1}^{n} \alpha_i y^i K(\mathbf{x}^i, \mathbf{x}) + b \tag{2}$$

The prediction function of general SVM:

$$f(\phi(\mathbf{x}); \mathbf{w}, b) = \begin{cases} 1 & iff\ (2) > 0 \\ -1 & otherwise \end{cases}$$

LOOCV error:

$$\frac{1}{n}\sum_{i=1}^{n}\delta(y^i, f(\phi(\mathbf{x}); \mathbf{w}^{-i}, b^{-i}))$$

where the superscript $-i$ denotes the parameters we found by removing the ith training example, and δ is an indicator function. Consider two cases:

1. Removing a support vector data point. The ith data point lies on the margin, and might be classified wrong. Because for such points, $\alpha_i > 0$, and might affect equation (2).

2. Removing a non-support vector data point. The ith data point lies outside the margin, and will be classified correctly for sure. Because for such points, $\alpha_i = 0$, and will not affect equation (2)

For case NO. 1, let's consider the worst case, that all m support vectors are classified wrong. This worst case leads to the upper bound of the LOOCV error $= \frac{m}{n}$

2 Implementation of SVMs

1.