Machine Learning, 2024 Fall
Homework 4

Notice

Due 23:59 (CST), Dec 26, 2024
Plagiarizer will get 0 points.
LaTeXis highly recommended. Otherwise you should write as legibly as possible.

1 Support Vector Machine [30pts]

1. Recall the hard-margin SVM objective:

$$\text{minimize } \frac{1}{2}\|\mathbf{w}\|^2$$
$$\text{subject to } y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1 \quad \forall i$$

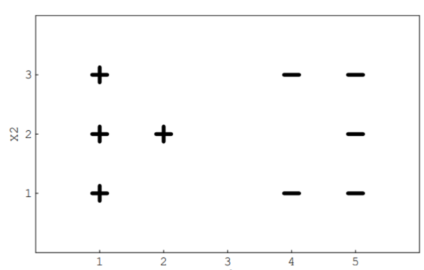
The constraints specify that the (functional) margin of each example is at least 1. If we change the constraint to require the margin to be at least c ($c > 0$), i.e., solving:

$$\text{minimize } \frac{1}{2}\|\mathbf{w}\|^2$$
$$\text{subject to } y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq c \quad \forall i$$

(1) Would it change the separating hyperplane? Why or why not?

(2) Let $\mathbf{w}^*$ be the solution of the original hard-margin SVM, and $\mathbf{w}_0$ be the solution of the modified problem with margin at least c. Write an expression for $\mathbf{w}_0$ in terms of $\mathbf{w}^*$.

2. Suppose we are using a linear SVM (i.e., no kernel), with some large C value, and are given the following data set.



(1) Draw the decision boundary of linear SVM. Give a brief explanation.

(2) In the above image, circle the points such that removing that example from the training set and retraining SVM, we would get a different decision boundary than training on the full sample. You need to offer a brief explanation.

2 Kernel Function [20pts]

(a) Let $k_1(u, v)$ be a valid kernel. Consider the new kernel function $k(u, v) = \exp(k_1(u, v))$, where $\exp(x)$ is the standard exponential function. Prove that $k(u, v)$ is also a valid kernel, i.e., show that it is positive semi-definite.

(b) Let $K_1(x, z)$ and $K_2(x, z)$ be valid kernels. Proving that for non-negative constants c_1 and c_2, $K_0(x, z) = c_1 K_1(x, z) + c_2 K_2(x, z)$ is a valid kernel function.

3 Support Vector Machine with Kernel [20pts]

Suppose we use a Support Vector Machine (SVM) with a custom kernel defined as:

$$K(x, x') = \begin{cases} 1 & \text{if } x = x', \\ -1 & \text{if } x \neq x'. \end{cases}$$

This corresponds to mapping each x to a vector $\psi(x)$ in some high-dimensional space (that need not be specified) so that

$$K(x, x') = \psi(x)^T \psi(x').$$

As in the original setup, we are given m training samples $(x_1, y_1), \ldots, (x_m, y_m)$, where $y_i \in \{-1, +1\}$, and all the data points x_i are distinct (i.e., $x_i \neq x_j$ for $i \neq j$).

Based on the standard SVM optimization problem, derive the expression of α_i when using the kernel defined above. Recall that the weight vector w used in SVMs has the form

$$\mathbf{w} = \sum_i \alpha_i y_i \psi(x_i)$$

4 K-Means [30pts]

Recalling the K-means, we iteratively find the cluster centers μ_t^k and upadte the class C_t^k for all data. Given a clusters number K, our goal is to minimize SSE

$$SSE = \sum_{k=1}^{K} \sum_{i \in C_t^k} \left\| x_i - \mu_t^k \right\|_2^2$$

(1) **Please prove that the K-means algorithm converges.**

(2) **Implement the K-means algorithm on the dataset we provide in Zip.** Your answer should include: embedded code, comment on your code and visual screenshot of your clustering results of K=2,5,10. Hint: Implement the K-means algorithm by hand (Dont use the sklearn implementation)