Machine Learning, 2024 Fall
Quiz 9

1 True or False [24pts]

1. The KNN algorithm is an instance-based learning method that doesn't require a training phase but instead stores all training data for prediction.
   solution: True

2. KNN is highly sensitive to outliers (i.e., data points far from the general distribution).
   solution: False

3. SVM classifies data by finding the hyperplane that maximizes the margin between classes.
   solution: True

4. In SVM, support vectors are the points that are farthest away from the decision hyperplane.
   solution: False

5. In SVM's hard margin classification, the data must be linearly inseparable.
   solution: False

6. Clustering algorithms can automatically determine the number of clusters.
   solution: False

7. K-means clustering requires the user to specify the number of clusters k in advance.
   solution: True

8. The optimal value of k in K-means clustering can be selected using cross-validation by minimizing cost function:
   $$C(z, \mu) = \sum ||x_i - \mu_{z_i}||^2$$
   solution: False

2 Conception [20pts]

(1) Why can both very small and very large k values be problematic in KNN? Based on this, explain how k influences the effective number of parameters in the KNN algorithm (i.e. How does the effective number of parameters change when the value of k increases or decreases?).

solution: (a) Small k: When $k = 1$, the model becomes highly sensitive to noise and outliers, leading to overfitting with a very complex decision boundary. This results in high variance and poor generalization. Large k: When k is too large, the model smooths the decision boundary, causing it to miss important patterns in the data, leading to underfitting with high bias.

(b) The effective number of parameters in KNN is inversely proportional to K (roughly N / k). When K is small (e.g., $K = 1$), the model is highly flexible, sensitive to individual points, and prone to overfitting, with high variance and low bias. The effective number of parameters is large because each data point has a strong influence on the model. When K is large, the model becomes smoother and less sensitive to individual points, reducing variance but increasing bias. The effective number of parameters decreases as the influence of each data point diminishes.

(2) Briefly explain how SVM uses the kernel method to handle non-linear classification.

solution: The kernel method allows SVM to implicitly map the data into higher-dimensional space where it is easier to find a hyperplane that separates the classes. Common kernels include linear, polynomial, and radial basis function (RBF). This avoids explicitly transforming the data and allows efficient computation.

3 K Nearest Neighbor [20pts]

The table below provides a training dataset containing six observations, three predictors, and one qualitative response variable:

Obs.	**X1**	**X2**	**X3**	**Y**
1	0	3	0	Red
2	2	0	0	Red
3	0	1	3	Red
4	0	1	2	Green
5	-1	0	1	Green
6	1	1	1	Red

(a) Compute the Euclidean distance between each observation and the test point $(0, 0, 0)$. **solution:** For a test point $(0, 0, 0)$, the Euclidean distance to each observation is calculated as follows:

$$d = \sqrt{(X_1 - 0)^2 + (X_2 - 0)^2 + (X_3 - 0)^2}$$

solution:

Observation	**Distance to Test Point**
1	$\sqrt{0^2 + 3^2 + 0^2} = \sqrt{9} = 3$
2	$\sqrt{2^2 + 0^2 + 0^2} = \sqrt{4} = 2$
3	$\sqrt{0^2 + 1^2 + 3^2} = \sqrt{10}$
4	$\sqrt{0^2 + 1^2 + 2^2} = \sqrt{5}$
5	$\sqrt{(-1)^2 + 0^2 + 1^2} = \sqrt{2}$
6	$\sqrt{1^2 + 1^2 + 1^2} = \sqrt{3}$

(b) Predictions the class of test point $(0, 0, 0)$ for K is 1, 3, and 6.

solution: $K = 1$: Look at the closest neighbor, which has the smallest distance. Observation 5 ($\sqrt{2}$) is the closest, with the label **Green**. Prediction: Green.

$K = 3$: Look at the 3 nearest neighbors: Observations 2 (**Red**), 5 (**Green**), and 6 (**Red**). Perform majority voting: 2 Reds and 1 Green. Prediction: Red.

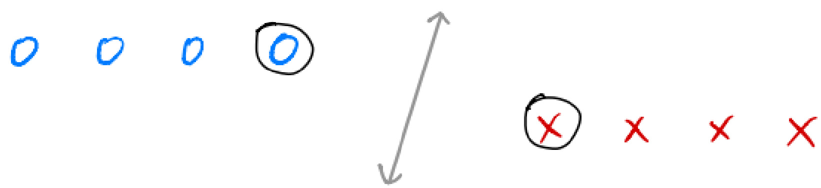
$K = 6$: Use all 6 observations for prediction. Perform majority voting: 4 Reds and 2 Greens. Prediction: Red.

4 Support Vector Machine [20pts]

Consider the following dataset: where circles are positive examples and crosses are negative examples.

(i) Draw the decision boundary given by a linear (hard-margin) SVM trained on this dataset, and circle the support vectors

solution:



(ii) Add a single example to the training set such that all examples would be support vectors.

solution:

5 K Means [16pts]

Given the following points in (x, y)-coordinates:

$$(3,4), (2,5), (8,11), (3,2), (6,3), (9,12)$$

Let the distance between two points (x_1, y_1) and (x_2, y_2) be defined by the custom distance metric:

$$d = 2|x_1 - x_2| + 3|y_1 - y_2|$$

(1) Assume the initial cluster centers are at $(3,3)$ and $(9,9)$. Perform 2-means clustering and determine the resulting clusters after one iteration.

solution:

Observation	x	y
1	3	4
2	2	5
3	8	11
4	3	2
5	6	3
6	9	12

$$\text{Centers:} \quad \text{Center } 1 = (3,3), \quad \text{Center } 2 = (9,9)$$

$$\text{For observation 1 } (3,4): \quad d_1 = 3, \quad d_2 = 27 \quad \Rightarrow \text{Assigned to Cluster 1}$$

$$\text{For observation 2 } (2,5): \quad d_1 = 8, \quad d_2 = 26 \quad \Rightarrow \text{Assigned to Cluster 1}$$

$$\text{For observation 3 } (8,11): \quad d_1 = 34, \quad d_2 = 8 \quad \Rightarrow \text{Assigned to Cluster 2}$$

$$\text{For observation 4 } (3,2): \quad d_1 = 3, \quad d_2 = 33 \quad \Rightarrow \text{Assigned to Cluster 1}$$

$$\text{For observation 5 } (6,3): \quad d_1 = 6, \quad d_2 = 24 \quad \Rightarrow \text{Assigned to Cluster 1}$$

$$\text{For observation 6 } (9,12): \quad d_1 = 39, \quad d_2 = 9 \quad \Rightarrow \text{Assigned to Cluster 2}$$

$$\text{New Centers:} \quad \text{Cluster 1 center} = (3.5, 3.5), \quad \text{Cluster 2 center} = (8.5, 11.5)$$