Question Examples for the lecture part of the final test

PETE-219

Q.1

Q.1-1

The unsupervised learning deals with the labeled data set. So, we start from a dataset that we know the answer.

(True/False)

Q.1-2

Support Vector Machine basically determines a decision boundary that can maximize the margin between classes.

(True/False)

Q.1-3

The K-means clustering method is one of the unsupervised machine learning techniques. It requires specifying the number of clusters before the training.

(True/False)

Q.1-4

Principal Component Analysis (PCA) is taken to reduce high dimensional space to low dimension in order to avoid overfitting or to show better visualization. For example, PCA performs eigenvalue decomposition and takes some of eigenvectors that correspond to the lower (smaller) values of eigenvalues.

(True/False)

Q.1-5

The neural networks (also called artificial neural networks) take one or more hidden layers between input and output layers, which can allow complex nonlinear calculation. Backpropagation is the learning process to find an optimal model with the weights that can minimize the loss function from the output errors.

(True/False)

Q.2 (Logistic regression) simple calculation

For binary classification $y \in \{0,1\}$, we obtain the following weights $(\theta_0, \theta_1, \theta_2, \theta_3)$ after the training of Logistic Regression, where each sample has three features (x_1, x_2, x_3).

$\theta_0 = -10,\ \theta_1 = -10,\ \theta_2 = -5,\ \theta_3 = 15$.

Then, by using the sigmoid function, prediction of y for a given sample that has $x_1 = 1,\ x_2 = 0,\ x_2 = 2$.

Specifically,

$$\begin{cases} y=1 & \text{when} \quad \phi(z)>0.5 \\ y=0 & \text{when} \quad \phi(z)<0.5 \end{cases},$$

$$\phi(z)=\frac{1}{1+\exp(-z)},\ z=\theta_0+\theta_1x_1+\theta_2x_2+\theta_3x_3$$

What is the predicted value of y for the given sample when there is no regularization?

Following Q.2-1, what is the predicted value of y for the given sample when an extremely large value is applied to the regularization term (e.g, $\lambda = 10^6$)?

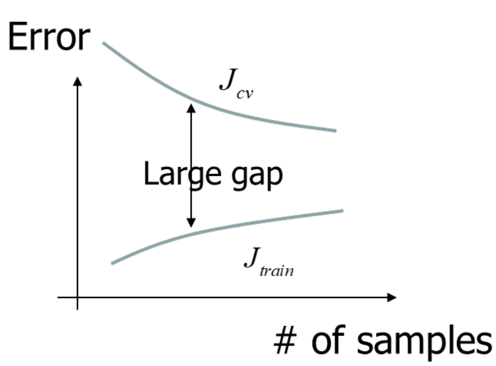
Q.3 (F1 score)

A confusion matrix for binary classification is given as follows from the total 130 samples.

	Predicted: Negative	Predicted: Positive
Actual: Negative	80	12
Actual: Positive	14	44

Find and answer (Q.3-1) the precision, (Q.3-2) recall, and (Q.3-3) F1 score, respectively.

Q.4 (overfitting, optimal, underfitting)



J_{cv} : cost function value (error) for cross validation, J_{train} : cost function value (error) for training

Which of the following cases can yield the above figure?

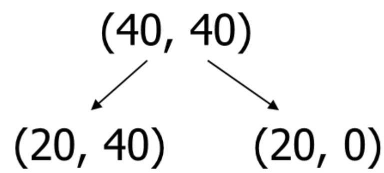
(a) overfitting (high variance)

(b) optimal

(c) underfitting (high biased)

Q.5 (DT-IG calculation)

Suppose that the data samples are divided into the following tree structure after using a binary decision tree training.



The information gain for the binary decision tree can be obtained by

$IG(D_p) = I_G(D_p) - \frac{N_{left}}{N_p} I_G(D_{left}) - \frac{N_{right}}{N_p} I_G(D_{right})$, where the subscript *p* denotes the parent node.

Gini impurity $I_G(t)$ at a node *t* is defined as

$I_G(t) = \sum_{i=0}^{1} p(i \mid t)(1 - p(i \mid t)) = 1 - (p(i=0 \mid t)^2 + p(i=1 \mid t)^2)$, where $p(i \mid t)$ is the probability at a node *t* that a sample belongs to Class *i.*

Calculate (Q.5-1) $I_G(D_{left})$, (Q.5-2) $I_G(D_{right})$, (Q.5-3) $IG(D_p)$.

Q.5

We use the majority voting based on probability for binary classification. For binary classification $y \in \{0,1\}$, suppose that we use 3 classifiers (i.e., C1, C2, C3), Each classifier has the following probability (p)and weight (w) of the majority voting for prediction of y of a given sample. Prediction is based on Logistic Regression just like Q2.1.

$$C1: p_{c1}(y=0) = 0.9, \quad w_{c1} = 0.2$$
$$C2: p_{c2}(y=0) = 0.7, \quad w_{c2} = 0.2$$
$$C2: p_{c3}(y=0) = 0.4, \quad w_{c3} = 0.6$$

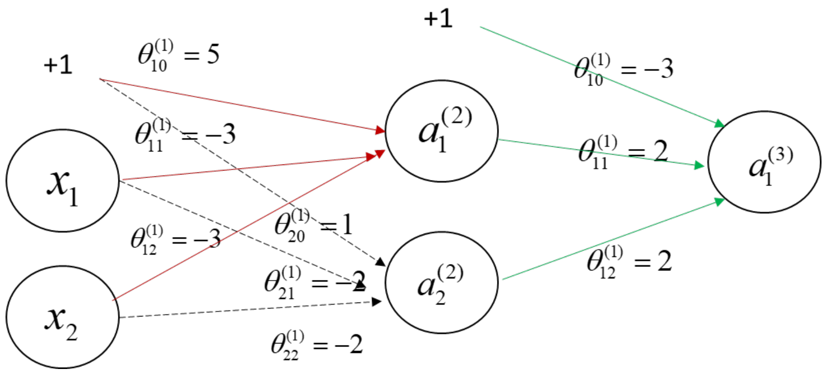
What is the predicted value of y for the given sample?

Q.6

Suppose that we have the following neural network structure. The first hidden layer takes the ReLU function as an activation function (ie, for $a_1^{(2)}, a_2^{(2)}$) while the output layer has the sigmoid function (ie, $a_1^{(3)}$). Concretely,

ReLU: Rectified linear unit $\phi(z) \begin{cases} z & \text{when } z \geq 0 \\ 0 & \text{when } z < 0 \end{cases}$

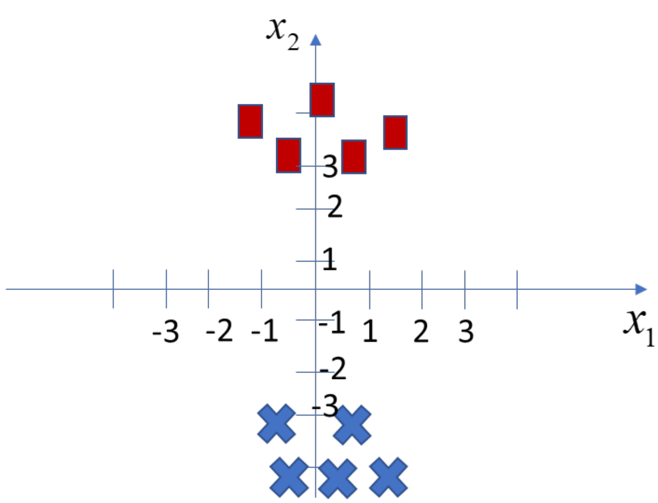
Sigmoid function, $\phi(z) = \frac{1}{1 + \exp(-z)}$



For a sample that has ($x_1 = -2, x_2 = 1$), calculate and answer (Q.6-1) $a_1^{(2)}$ and (Q.6-2) $a_2^{(2)}$.

Q.7

$$\underbrace{\text{Min}}_{\theta} \frac{1}{2} \underbrace{\left(\theta_1^2 + \theta_2^2\right)}_{\|\boldsymbol{\theta}\|^2} \quad \text{s.t.} \begin{pmatrix} z^i \geq 1 \leftarrow y^i = 1 \\ z^i \leq -1 \leftarrow y^i = 0 \end{pmatrix}, \ \boldsymbol{\theta} = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}, \ z^i = \boldsymbol{\theta}^T \mathbf{x}^{(i)} = \theta_1 x_1^{(i)} + \theta_2 x_2^{(i)} (= \|\boldsymbol{\theta}\| p^{(i)}),$$ where the superscript i denotes the sample index. $p^{(i)}$ is the (signed-positive or negative) projection of $\mathbf{x}^{(i)}$. The figure below shows distribution of features of the samples.



At the optimal value of $\boldsymbol{\theta}$, what is $\|\boldsymbol{\theta}\|$?

Q.8 PCA

For a given covariance matrix, please find eigen vectors and eigen values, first. Then, find W followed by $\mathbf{xW} = \mathbf{z}$. Refer to the slides of PCA.