Artificial Intelligent and Machine Learning

Time: 1h30

Students are allowed to use their personal calculator

Exercise 1 (7points)

1a. Explain the idea of gradient descent algorithm, define the cost function and the rule of updating coefficient of linear regression.

1b. Explain why the cost function of logistic regression can be estimated by the following formulation:

$$E(\theta) = -\frac{1}{m}\sum_{i=1}^{m}[y^{(i)} \log h_\theta(x^{(i)}) + (1 - y^{(i)}) \log(1 - h_\theta(x^{(i)})]$$

1c. Assuming at a certain iteration of batch-gradient descent, we reach the global **maximum** value of the cost function ($\frac{\partial}{\partial_\theta} E(\theta)$ is perfectly 0). What will happen next and why?

- The algorithm stays at the global maximum
- The algorithm updates the coefficient toward the minimum value of cost function

Is the answer remaining the same if we are using mini-batch gradient descent or stochastic-gradient descent?

Exercise 2 (3 points)

What characteristic of covariance matrix make it useful in PCA implementation? Why?

Exercise 3 (5 points): Given the following data

X_1	X_2
1	1
4	1
3	2
2	4
1	3
3	1

Using k-means algorithm to cluster the training data into 2 groups with the initial value of 2 group are

Group 1: (1,1)

Group 2: (4,2)

What is the centroid values of two group after 1st and 2nd iteration?

Exercise 4 (5 points)

Given the training data for a classification problem, to identify if a person is male/female base on his/her weight and high. Stating at $\theta_0 = \theta_1 = \theta_2 = 0$, and learning rate α=0.0001. Calculate the coefficients after the first iteration with batch-gradient descent.

Training Example	Height (cm)	Weight (kg)	Male/Female
1	172	68	Male
2	163	52	Female
3	158	50	Female
4	180	75	Male