Problem 1

Given a data set $D=\{(\boldsymbol{x}_i, y_i)\}_{i=1}^{m}$, where $\boldsymbol{x}_i \in \mathbb{R}^d$ and $y_i \in \mathbb{R}$. We want to use an unregularized least squares regression model to best fit this data set. This can be formulated as the following optimization problem:

$$\min_{\boldsymbol{w}\in\mathbb{R}^d, b\in\mathbb{R}} \ell(\boldsymbol{w}, b) := \frac{1}{2}\sum_{i=1}^{m}(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i + b - y_i)^2. \tag{1-1}$$

Try to answer the following questions:

(1) Is the optimal parameter $(\boldsymbol{w}^*, b^*)$ unique? If not, please give the condition for guaranteeing the uniqueness of $(\boldsymbol{w}^*, b^*)$.

(2) The data set D is shown in the Table 1, where each sample has 3 dimensions (d_1, d_2, d_3). Please calculate the optimal parameter $(\boldsymbol{w}^*, b^*)$.

Table 1 Training set for linear regression

d_1	2	9	8	8	2	8	4	1	3	5
d_2	9	3	3	8	1	4	3	8	3	3
d_3	5	4	2	1	6	7	8	2	6	7
y	8	10	6	7	3	4	12	7	5	4

Problem 2

In a binary classification problem, each instance $\boldsymbol{x}_i \in \mathbb{R}^d$ in a data set $D=\{(\boldsymbol{x}_i, y_i)\}_{i=1}^{m}$ has a label $y_i \in \{0,1\}$. We have already known the logistic regression model Eq.(2-1) is a powerful tool to handle this kind of problem.

$$y = \frac{1}{1+e^{-(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}+b)}} \tag{2-1}$$

To simplify this problem, we assume that $\boldsymbol{\beta} = (\boldsymbol{w}; b)$, $\hat{\boldsymbol{x}} = (\boldsymbol{x}; 1)$. Since its negative log-likelihood function Eq.(2-2) is convex, we can optimize it efficiently with Gradient Descent method, Newton Method, and so on.

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^{m}\left(-y_i\boldsymbol{\beta}^{\mathrm{T}}\hat{\boldsymbol{x}}_i + \ln\left(1+e^{\boldsymbol{\beta}^{\mathrm{T}}\hat{\boldsymbol{x}}_i}\right)\right) \tag{2-2}$$

(1) Prove the Eq.(2-2) is convex.

(2) Suppose we are facing a K-class classification problem instead of a binary classification problem, where $y_i \in \{1, 2, \cdots, K\}$. Please expand the logistic regression model Eq.(2-1) to a multi-class version and give the log-likelihood function of this multi-class logistic regression model.

Problem 3

In a binary classification problem, given the true label y of the sample and the predicted values y_{C_1}, y_{C_2} of the two classifiers C_1, C_2, calculate the relevant performance measures.

Table 2 True label and predicted values of two classifiers

y	1	0	1	1	1	0	0	1
y_{C_1}	0.62	0.39	0.18	0.72	0.45	0.01	0.32	0.93
y_{C_2}	0.36	0.12	0.82	0.89	0.17	0.75	0.36	0.97

(1) AUC

(2) Confusion Matrix (threshold=0.4 and 0.9 for C_1 and C_2 respectively)

(3) $F1$-Score (threshold=0.4 and 0.9 for C_1 and C_2 respectively)

Problem 4

Suppose you have a regression problem where the training data has 1000 samples and each sample has 20 features. Now you fit the training data with a polynomial regression model of degree d and use mean squared error (MSE) as the loss function. Try to answer the following questions:

(1) If $d = 1$, is the model prone to overfitting? Why?

(2) If $d = 20$, is the model prone to overfitting? Why?

(3) How to choose an appropriate d value to balance the bias and variance of the model?