ECE449 Assignment Conceptual Questions

XieTian 3180111631

Problem 1,

a) ① supervised learning as a multi-classifier problem on books. The training data is a set of multi-dimensional personal interest feature vectors.

② reinforcement learning, where the system gets positive feedback when the shopper buys the recommended book and negative otherwise. The training data is a set of multi-dimensional personal interest feature vectors.

b) reinforcement learning, where the training data is the in-time score of the game. The model gets positive feedback if the score in game increases and negative feedback otherwise.

c) supervised learning, where the training data is the set of (movie, types) tuples. This is a multi-classification problem.

d) unsupervised learning, where the training data is the set of music pieces. The model discovers the inner structures of these good music pieces and generates similar new musics.

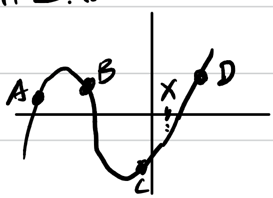
e) supervised learning, where the training data is the set of (max debt allowed, ±1). ±1 represent bankrupt or not. This is a binary classification problem.

Problem 2,

Linear regression mostly is used for fitting existing data curves and predict the continuous scalar values.

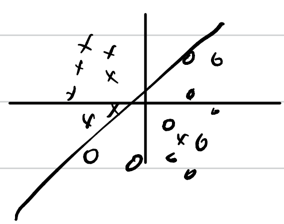
Logistic regression mostly is used for classification, which predicts discrete values 0, 1, 2 ...

Linear regression



fitting the model curve based on known A, B, C, D, predicting the unknown value on x

Logistic regression



learning the boundary between "x" and "o" classify an unknown point

Problem 3,

From lecture we know. $\hat{w} = (H^T H)^{-1} H^T b$

Here we have $H = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix}$, $b = \begin{pmatrix} 2 \\ 1 \\ 2 \end{pmatrix}$

With trivial computation we obtain $\hat{w} = (\frac{5}{3}, 0)^T$

problem 4,

$$H = \begin{pmatrix} 1 & -2 & (-2)^2 \\ 1 & -1 & (-1)^2 \\ 1 & 1 & 1^2 \\ 1 & 2 & 2^2 \end{pmatrix} = \begin{pmatrix} 1 & -2 & 4 \\ 1 & -1 & 1 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{pmatrix}$$

$y = (1, 0, 0, 2)^T$

problem 5,

① First define a loss function $J(w)$ to evaluate the performance of the trained model.

② while ! convergence:

$w \mathrel{-}= \alpha \cdot \frac{\partial}{\partial w} J$ (α is a parameter, learning rate)

③ return w

in the ②, the gradient asent algorithm uses the following code to update.

while ! convergence:

$$w \mathrel{+}= \alpha \cdot \frac{\partial}{\partial w} J$$ (α has the same meaning)

Since gradient desent algorithm seeks the minimal value, in the contrary, the gradient asent algorithm seeks the maximal value.

Problem 6,

advantages:

large learning rate: converges faster, less iteration needed

small learning rate: extra precision, more likely to find minimum

disadvantages:

large learning rate: may skip the minimum in one update and fail to converge.

small learning rate: takes too much time and computation

Problem 7,

$$\text{entropy} = -\sum_{i=1}^{4} P_i \log_2 P_i = -\tfrac{1}{4}\log_2 \tfrac{1}{4} \times 4 = 2$$

Problem 8,

$$H(X|Y) = \sum_{y\in Y} P(y)\, H(X|Y=y) = -\sum_{y\in Y} P(y) \sum_{x\in X} P_{X|Y}(x|y) \log_2 P_{X|Y}(x|y)$$

$$= -\tfrac{1}{4}\left(\tfrac{1}{2}\log_2\tfrac{1}{2} + \tfrac{1}{4}\log_2\tfrac{1}{4} + \tfrac{1}{8}\log_2\tfrac{1}{8} + \tfrac{1}{8}\log_2\tfrac{1}{8}\right) - \tfrac{1}{4}\left(\tfrac{1}{4}\log_2\tfrac{1}{4}\times 4\right)$$

$$-\tfrac{3}{8}\left(\tfrac{1}{3}\log_2\tfrac{1}{3}\times 2 + \tfrac{1}{6}\log_2\tfrac{1}{6}\times 2\right) - \tfrac{1}{8}\, 1\log_2 1$$

$$= 1.657$$

similarly $H(Y|X) = -\sum_{x\in X} P(x) \sum_{y\in Y} P_{Y|X}(y|x) \log_2 P_{Y|X}(y|x)$

$$= -\tfrac{7}{16}\left(\tfrac{2}{7}\log_2\tfrac{2}{7}\times 3 + \tfrac{1}{7}\log_2\tfrac{1}{7}\right) - \tfrac{1}{4}\left(\tfrac{1}{4}\log_2\tfrac{1}{4}\times 2 + \tfrac{1}{2}\log_2\tfrac{1}{2}\right)$$

$$-\tfrac{5}{32}\left(\tfrac{2}{5}\log_2\tfrac{2}{5}\times 2 + \tfrac{1}{5}\log_2\tfrac{1}{5}\right) - \tfrac{5}{32}\left(\tfrac{2}{5}\log_2\tfrac{2}{5}\times 2 + \tfrac{1}{5}\log_2\tfrac{1}{5}\right)$$

$$= 1.704$$

$$I(X;Y) = H(Y) - H(Y|X) = -\sum_{y\in Y} P(y)\log_2 P(y) - H(Y|X)$$

$$= -\left(\tfrac{1}{4}\log_2\tfrac{1}{4} + \tfrac{1}{4}\log_2\tfrac{1}{4} + \tfrac{3}{8}\log_2\tfrac{3}{8} + \tfrac{1}{8}\log_2\tfrac{1}{8}\right) - 1.704$$

$$= 0.202$$

Problem 9.

$$X^T X = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_N \end{pmatrix} \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_N \end{pmatrix} = \begin{pmatrix} N & \sum_{j=1}^{N} x_j \\ \sum_{j=1}^{N} x_j & \sum_{j=1}^{N} x_j^2 \end{pmatrix}$$

then $$\vec{x}_i (X^T X)^{-1} \vec{x}_i^T = (1 \ x_i)\left(N \Sigma x_j^2 - (\Sigma x_j)^2\right)^{-1} \begin{pmatrix} \Sigma x_j^2 & -\Sigma x_j \\ -\Sigma x_j & N \end{pmatrix} \begin{pmatrix} 1 \\ x_i \end{pmatrix}$$

$$= \left(N \Sigma x_j^2 - (\Sigma x_j)^2\right)^{-1} \left(\Sigma x_j^2 - 2x_i \Sigma x_j + N x_i^2\right)$$

let $\sum_{j=1}^{N} x_j = \bar{x} N$, now we have:

$$\vec{x}_i (x^T x)^{-1} \vec{x}_i^T = \frac{\sum_{j=1}^{N} x_j^2 - 2N\bar{x} x_i + N x_i^2}{N \sum_{j=1}^{N} x_j^2 - N^2 \bar{x}^2}$$

$$= \frac{1}{N} + \frac{\bar{x}^2 - 2\bar{x} x_i + x_i^2}{\sum_{j=1}^{N} x_j^2 - N\bar{x}^2}$$

$$= \frac{1}{N} + \frac{(\bar{x} - x_i)^2}{\sum_{j=1}^{N} x_j^2 - N\bar{x}^2} = \frac{1}{N} + \frac{(\bar{x} - x_i)^2}{\sum_{j=1}^{N} (x_j^2 - \bar{x}^2)}$$

therefore, we prove $\vec{x}_i (x^T x) \vec{x}_i^T = h_i$

$h_i = \vec{x}_i (x^T x)^{-1} \vec{x}_i^T$

let $H = \vec{x} (x^T x) \vec{x}^T = H$

$\beta = (X^T X)^{-1} X^T \vec{y}$ $\quad X\beta = \underbrace{X(X^T X)^{-1} X^T}\vec{y} = \vec{y}$

So $H\vec{y} = \vec{y}$