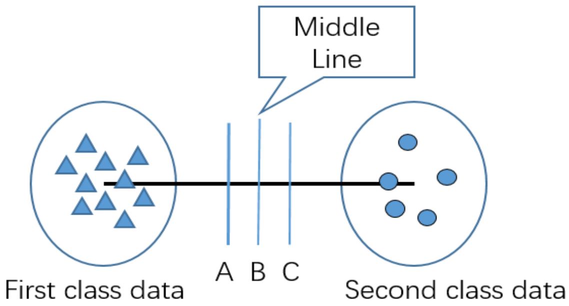
统计学习理论与方法期中考试题

Student ID______________ Name ________________Score ________

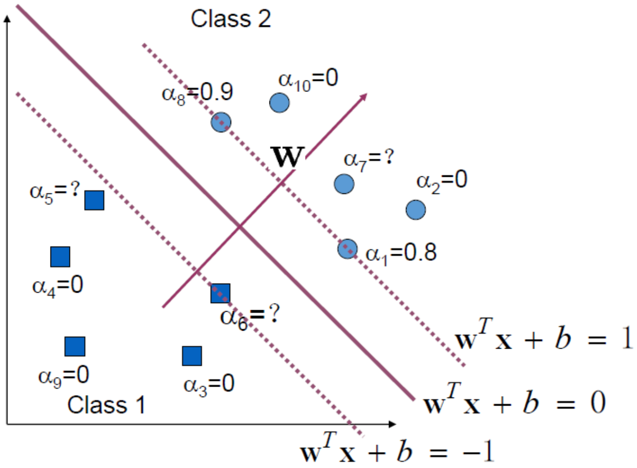
一、 (20 points) Check the following statements(True or false)：

1. The complexity of KNN (K-Nearest Neighbors) regression models is described by the model parameter K. The larger K is, the higher complexity of KNN model is.
2. For the linear regression model $RSS(\beta)=\sum_{k=1}^{N}(y_k-\sum_{j=0}^{p}\beta_j x_k^j)^2$, where p is the order of polynomial functions. The variance of the linear regression estimator will increase as the order p increases.
3. The Vapnik-Chervonenkis (VC) dimension of a 5-dimensional linear function family $\left\{ f(x,\alpha)=\alpha_0+\sum_{i=1}^{5}\alpha_i x_i,\ \ x\in R^5 \right\}$ is 5.
4. For two class linear classification problem， the number of first class training sample is N_1, the number of second class training sample is N_2, $N_1 > N_2$. Both classes follow the Gaussian distribution with the same covariance. Please check which straight line (A, B, C) is the most likely correct decision boundary of LDA as showed in the following figure.



二、(30 points) Answer the following questions

1. Formulate the EM algorithm for the two Gaussian mixture model by introducing a latent variable, in order to give an explicit solution to the maximum step. Please describe the EM algorithm for the two Gaussian mixture model.
2. Consider linearly separable support vector machine problem, $\alpha_1,\alpha_6,\alpha_8$ are the Lagrange multiplier coefficients of support vectors. Please calculate the coefficient values for $\alpha_5,\alpha_6,\alpha_7$ according to the following figure's data.



三、（20 Points ）Consider a linear regression model $Y=[1,X^T]\beta+\varepsilon,\quad \beta=(\beta_0,\beta_1,\cdots,\beta_p)^T$, and the probability density function of noise ε is $N(0,\sigma_\varepsilon^2 I)$. Given training data $\{(x_i,y_i)\}_{i=1}^N$, $x_i\in R^p$, $y_i\in R^1$.

1. Assume $\hat{f}(X)$ is the regression function. Decompose the expected prediction error $Err(x_0)=E[(Y-\hat{f}(x_0))^2\mid X=x_0]$ into the model bias and variance.
2. Derive the regression solution $\hat{\beta}$ and calculate the variance of the regression solution var($\hat{\beta}$).
3. Calculate the variance of the regression estimator var($\hat{f}(x_0)$).

四、（15 points ）Assume the probability density function of each class data is Gaussian,

$$N\left(x\middle|\mu_k,\Sigma_k\right)=\frac{1}{(2\pi)^{p/2}\sqrt{|\Sigma_k|}}\exp\left\{-\frac{1}{2}(x-\mu_k)^T\Sigma_k^{-1}(x-\mu_k)\right\},\quad k=1,2,\ldots,K,\quad x\in R^p.$$

1. What is the class mask problem of linear classification by using linear regression method? How to solve the problem?
2. Describe the procedure of the reduced LDA.

五、（15 points ） Consider local linear regression problem

$$\min_\beta\sum_{i=1}^N K_\lambda(x_0,x_i)\left[y_i-[1\ x_i^T]\beta\right]^2$$

Assume that $\{(x_i,y_i)\}_{i=1}^N$ is the training data, $x_i\in R^p, y_i\in R^1$.

1. Please derive the regression solution for $\hat{\beta}$.
2. Denote the regression function as $\hat{f}(x_0)=\left[1\ \ x_0^T\right]\hat{\beta}(x_0)=\sum_{i=1}^N l_i(x_0)y_i$. Prove the following relation

$$\sum_{i=1}^N l_i(x_0)=1;\quad \sum_{i=1}^N x_i l_i(x_0)=x_0.$$