Problem 1: Answer the following questions（10 Points，5 points for each problem）

1. How to define the generalization error for regression models? Please provide three approaches that are able to improve the generalization performance for the regression models.
2. What is the model selection problem? Please provide two model selection criteria?

Problem 2: (15 Points) Consider KNN model for regression problem $\{(x_i, y_i)\}_{i=1}^{N}$ $N_K(x_0)$ is the set of K-closest samples to x_0. The KNN prediction model is given by

$$\hat{f}(x_0)=\frac{1}{k}\sum_{x_j \in N(x_0)} f(x_j),$$

1. Please derive the model variance $\text{var}[f(x_0)] = \sigma_\varepsilon^2 / k.$

2. Explain that why the KNN method is not suitable for a high dimensional regression problem.

Problem 3: (15 Points) Suppose that $\{(x_i, y_i)\}_{i=1}^{N}$ is N-training samples, $x_i \in R^p$。Prove that the estimated prediction function $\hat{f}(x_0) = \beta_0 + \beta^T x_0 = \bar{x}_0^T \bar{\beta}$, where $\bar{x}_0 = (1, x_0^T)^T$ to the linear regression model can be rewritten in the following form:

$$\hat{f}(x_0) = \sum_{i=1}^{N} l_i(x_0, X) y_i$$

where the weight $l_i(x_0, X)$ do not depend on the y_i, $X = (\bar{x}_1^T, \bar{x}_2^T, \cdots, \bar{x}_N^T)^T$ is the data matrix.

1. Please derive the detailed expression for $l_i(x_0, X)$.

2. Calculate the variance of the prediction function $\hat{f}(x_0)$ in term of $l_i(x_0, X)$

Problem 4 (15 points) Assume the probability density function of k-class data is Gaussian,

1. Derive the posterior $p(G=k \mid x)$ in terms of the prior $p(G=k)$ and $p(x|G=k)$.
2. Provide unbiased estimation for covariance $\hat{\Sigma}_k$ of k-*th* class data.
3. Describe the procedure of the reduced rank LDA.

Problem 5 (15 Points) Assume that statistical model $Y = f(X) + \varepsilon$, where ε follows the normal distribution $N(0,\sigma_\varepsilon^2)$. Given training data $\{(x_i, y_i)\}_{i=1}^{L}$, $x_i \in R^p, y_i \in R^1$. $\hat{y} = \hat{f}(X)$ is the regression function。 We define an in-sample error as

$$Err_{in} = \frac{1}{N}\sum_{i=1}^{N} E_{Y^{new}}\left[L(Y_i^{new}, \hat{f}(x_i))|\mathrm{T}\right],$$

a) If the loss function is the squared error, prove that the Optimism is given by

$$op = Err_{in} - E_y[\overline{err}] = \frac{2}{N}\sum_{i=1}^{N} Cov(\hat{y}_i, y_i),$$

where $\hat{y}_i = \hat{f}(x_i)$, $\overline{\text{err}}$ is the training error.

b) For linear regression model, prove the following relation

$$\sum_{i=1}^{N} Cov(\hat{y}_i, y_i) = p\sigma_\varepsilon^2.$$

Problem 6: Consider a linear regression problem

$$Y = f(x) + \varepsilon; \qquad \varepsilon \sim N(0,\sigma^2),\ \ x = (x^1, x^2, \cdots, x^p)^T$$

$$f(x) = x^T\beta = \sum_{i=1}^{p} \beta_i x^i \quad .$$

Assume that $D = \{(x_i, y_i)\}_{i=1}^{N}$ is N-training samples.

- Please define likelihood function $p(y\,|\,x,\beta)$ for the regression problem. Derive the solution of maximum likelihood problem.
- Suppose that the parameter prior follows Gaussian distribution $\beta \propto N(0, \tau\,\Sigma)$, define its posterior distribution $p(\beta|y, x)$, and prove that the posterior is also Gaussian. Derive the mean and covariance for the posterior distribution.