Machine Learning and Applications

Assignment 6

1. Let $\mathcal{D} = \{(x_1, y_1), \dots, (x_N, y_N)\}$ be a given training set. Consider in-sample prediction error for the squared-error loss:

$$\mathrm{Err}_{in} = \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}_{Y^0}(Y_i^0 - \hat{f}(x_i))^2$$

and the training error $\overline{\mathrm{err}}$:

$$\overline{\mathrm{err}} = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{f}(x_i))^2$$

The notation Y^0 indicates that we observe new response values at each of the training points $x_i, i = 1, 2, \dots, N$, i.e. the response value y_i at each point x_i is considered as random variable. Let us define *optimism* as the difference between Err_{in} and the training error $\overline{\mathrm{err}}$:

$$\mathrm{op} \equiv \mathrm{Err}_{in} - \overline{\mathrm{err}}.$$

Add and subtract $f(x_i)$ and $\mathbb{E}\hat{f}(x_i)$ in each expression and expand. Hence establish that the average optimism in the training error is

$$\frac{2}{N} \sum_{i=1}^{N} \mathrm{Cov}(\hat{y}_i, y_i)$$

2. Let us consider the following model: $y = \beta^t \mathbf{x} + \epsilon$, where ϵ is a random noise with $\mathbb{E}\epsilon = 0, \mathbb{V}\epsilon = \sigma_\epsilon^2$. For a linear regression $\hat{\mathbf{y}} = \mathbf{S}\mathbf{y}$, show that

$$\sum_{i=1}^{N} \mathrm{Cov}(\hat{y}_i, y_i) = \mathrm{trace}(\mathbf{S})\sigma_\epsilon^2,$$

where $\mathbf{y} = (y_1, \dots, y_N)$ — vector of response values from the training set $\mathcal{D} = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_N, y_N)\}$, $\hat{\mathbf{y}} = (\hat{y}_1, \dots, \hat{y}_N)$ is a vector of predictions.