MSPE & Irreducible Error

Suppose that

$$Y_i = g(\vec{x}_i) + E_i,$$

where the E_i are iid r.v's having mean 0. Letting $\hat{g}_{m,n}$ denote an estimator of g based on

$$(\vec{x}_1, Y_1), (\vec{x}_2, Y_2), \ldots, (\vec{x}_n, Y_n),$$

and a specific method-model combination, m, suppose that it is desired to predict a value for Y_{n+1} that will be associated with a known vector, $\vec{x}_{n+1}$. (Note: It is assumed that $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_n$ are all known.) If it is assumed that the pdf of the E_i is symmetric and unimodal (assuming its maximum value at 0), then the unknown value, $g(\vec{x}_{n+1})$, would be a good prediction for Y_{n+1}. In practice, one can use $\hat{g}_{m,n}(x_{n+1})$ (viewing it as a r.v. — a fn of the r.v.s $Y_1, Y_2, \ldots, Y_n$) as a predictor of Y_{n+1}.

The *mean squared prediction error* (MSPE) is a measure which is commonly used to compare the accuracies of various predictors (although most of the time we have to be content with comparing estimated MSPE values). Letting σ^2 denote the variance of the "error term" dist'n (the dist'n of the E_i), we have

$$E([\hat{g}_{m,n}(\vec{x}_{n+1}) - Y_{n+1}]^2)$$

$$= E([\hat{g}_{m,n}(\vec{x}_{n+1}) - E(\hat{g}_{m,n}(\vec{x}_{n+1}))]^2)$$

$$+ E([E(\hat{g}_{m,n}(\vec{x}_{n+1})) - Y_{n+1}]^2)$$

$$= Var(\hat{g}_{m,n}(\vec{x}_{n+1})) + E([E(\hat{g}_{m,n}(\vec{x}_{n+1}) - g(\vec{x}_{n+1})]^2)$$

$$+ E([g(\vec{x}_{n+1}) - Y_{n+1}]^2)$$

$$= Var(\hat{g}_{m,n}(\vec{x}_{n+1})) + [bias(\hat{g}_{m,n}(\vec{x}_{n+1}))]^2 + \sigma^2.$$

(Note: To get each equality, I skipped showing some steps above which I will show on the board.) It can be seen that the mean squared prediction error associated with predicting Y_{n+1} cannot be less than σ^2. The variance of the error term dist'n, σ^2, is referred to as the *irreducible error*.

On the previous page, the mean squared prediction error was for predicting Y at a specific value, $\vec{x}$. If one just refers to the mean squared prediction error w/o specifying a particular $\vec{x}$, then the point at which the prediction is to be made should be taken to be random. For an estimate of this general MSPE we can use

$$\frac{1}{n_2} \sum_{i=n_1+1}^{n_1+n_2} (y_i - \hat{g}_{m,n_1}(\vec{x}_i))^2,$$

where $\hat{g}_{m,n_1}$ is the prediction rule fit from

$$(\vec{x}_1, y_1), \ (\vec{x}_2, y_2), \ \ldots, \ (\vec{x}_{n_1}, y_{n_1}),$$

and

$$(\vec{x}_{n_1+1}, y_{n_1+1}), \ \ldots, \ (x_{n_1+n_2}, y_{n_1+n_2})$$

are observations of an independent sample.