Chapter 2 : Statistical Learning

General form of an estimator is $Y = f(x) + e$. Here $f(x)$ is the systematic information that x provides about Y. We also know that $E(e) = 0$

Reducible vs Irreducible Error :
Expectation of difference in actual Y with respect to estimator predictions $\hat{Y}$ can be decomposed into two parts - variablity from error that can be reduced by improving the model vs variability from irreducible error. The irreducible error will always provide an upper bound on the accuracy of our prediction for Y. This bound is almost always unknown in practice. All models attempt to lower this reducible error.

$E(Y - \hat{Y})^2$ = $E(f(x) + e - f(\hat{x}))^2$

= $E((f(x) - f(\hat{x}))^2 + e^2 + 2e(f(x) - f(\hat{x})))$

Expanding and using $Var(e)$ = $E(e^2) - E(e)^2 = E(e^2)$

= $[f(x) - f(\hat{x})]^2 + Var(e) + Var(f(x) - f(\hat{x}))$

= $[f(x) - f(\hat{x})]^2 + Var(e)$

= $Variance(ReducibleError) + Variance(IrreducibleError)$

The above equation thus forms the basis of Bias-Variance trade-off identity in ML models.

$$E(y_0 - f(\hat{x_0}))^2 = Var(f(\hat{x_0})) + [Bias(f(\hat{x_0}))]^2 + Var(e)$$

To reduce difference between actual and predicted, low variance and low bias models are preferred.

Parametric Methods : Assumption made on the shape of $f(x)$ say, linear model assumes functional form of f as linear wrt all independent variables with parameters β_0, β_1, β_2 etc.

Non Parametric Methods : No assumption made on shape of curve, hence more flexible and generally able to avail a better fit. Difficult to solve though!
Eg. Thin Plate Splines, refers to a physical analogy involving the bending of a thin sheet of metal. Just as the metal has rigidity, the Thin Plate Splines fit resists bending, implying a penalty involving the smoothness of the fitted surface.
See more : Kriging Interpolation

Accuracy vs Interpretability : More accurate -> More complex modeling -> Hard to interpret

Underfit vs Overfit :
As a general rule, as we use more flexible methods, the variance will increase and the bias will decrease. The relative rate of change of these two quantities determines whether the test MSE increases or decreases. As we increase the flexibility of a class of methods, the bias tends to initially decrease faster than the variance increases. Consequently, the expected test MSE declines. How-

ever, at some point increasing flexibility has little impact on the bias but starts to significantly increase the variance. When this happens the test MSE increases and we get an overfit. Common way to avoid this is cross-validation.

KNN vs KMeans : Both work similarly using Minkowski distance metric (p=2) in supervised vs unsupervised setting respectively. Interestingly, owing to this Kmeans becomes highly susceptible to outliers, whereas KNN is not.