# Statistical Learning Notes

10/20/23
M. Ranganathan

**The Problem**: we have some system

$$Y = f(X) + \varepsilon$$

"response" variable — $Y$

$\varepsilon$ → error/noise term

"predictor" variables $X = [X_1, X_2, \dots, X_n]$

Ex: We want to predict future rainfall changes. Rainfall is the "response", and predictors are observable quantities that affect rainfall
e.g. temperature, sea level pressure, atmospheric circulations (ENSO, MJO, PDO, ...), wind speed, ...

However: We don't know $f$!

We want to use observations of the system to approximate $f$, for two goals:

[1] Prediction: We want to predict the response $Y$ of the system

$$\hat{Y} = \hat{f}(X)$$

Predicted $Y$ — $\hat{Y}$
$\hat{f}$ → our estimated form of $f$

For this purpose, $\hat{f}$ can be a black box.

[2] Inference: we want to infer the form of $f$ to understand how $Y$ changes as $X$ changes, in order to

→ identify which predictors $X_i$'s are most important to $Y$

→ understand the relationship between $Y$ and each $X_i$

To do both of these, we need to find $\hat{f}$.
How do we use data to do so?

We want to find some $\hat{f}$ such that

$$\hat{Y} = \hat{f}(X)$$

is close to the real response $Y$.

Say we have $p$ datapoints $(x_1, y_1) \dots (x_p, y_p)$

**Approach 1:** Linear Regression

Let's assume $f$ is linear:

$$\hat{f}(x) = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n$$

We've now simplified this problem to the estimation of $n+1$ parameters.

This is called a "parametric" method.

We can write this out as a function of our data:

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_p \end{bmatrix} = \beta_0 + \beta_1 \begin{bmatrix} X_{11} \\ \vdots \\ X_{1p} \end{bmatrix} + \dots + \beta_n \begin{bmatrix} X_{n1} \\ \vdots \\ X_{np} \end{bmatrix}$$

If $n+1 \leq p$ we can solve this problem by matrix algebra:

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_p \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & \cdots & X_{n1} \\ \vdots & \vdots & \ddots & \\ 1 & X_{1p} & \cdots & X_{np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_n \end{bmatrix}$$

**Approach 2:** Non-Parametric Methods

**Model Selection:**

Mean Squared Error:

$$MSE = \frac{1}{p} \sum_{i=1}^{p} (Y_i - \hat{Y}_i)^2$$

[See Matlab Live Script or iPython notebook for the remainder of the notes]