# MA506 Probability and Statistical Inference

# Lecture 7: Regression 2

In [1]:
```
import numpy as np
import matplotlib.pyplot as plt
```

# 1. Extending to a more general cases of Linear Regression

## 1.1 Instead of 4 samples, we have n samples

Modify the X, Y matrices to incorporate all the samples: $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}; \quad X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$$

And the optimal coefficients $\hat{\beta}$ will have the same expression as before (4) with updated X and Y. Hence still we will have

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

## 1.2 Instead of $y = \beta_0 + \beta_1 x$ we want : $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \ldots + \beta_m x^m$

Assuming we still have n data points, relative to the base case, now we need to update all matrices involved: $X$, $Y$ and $\beta$:

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}; \quad X = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^m \\ 1 & x_2 & x_2^2 & \cdots & x_2^m \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^m \end{bmatrix}; \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{bmatrix}$$

With these new matrices, the optimal coefficients (estimate of $\beta$: $\hat{\beta}$) will still have the same expression as (4)

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

## 1.3 Instead of $y = \beta_0 + \beta_1 x$ we want : $y = \beta_0 + \beta_1 sin(x) + \beta_2 cos(x)$

Assuming $n$ samples, again updating $X$, $Y$ and $\beta$ as follows

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}; \quad X = \begin{bmatrix} 1 & Sin(x_1) & Cos(x_1) \\ 1 & Sin(x_2) & Cos(x_2) \\ \vdots & \vdots & \vdots \\ 1 & Sin(x_n) & Cos(x_n) \end{bmatrix}; \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$$

Like before, with updated $X$, $Y$ and $\beta$, we still have:

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

## 1.4: Linear regression in multiple dimensions (Multiple Linear Regression)

Suppose we have n samples in d dimensions. Hence each sample has d coordinates. For instance, $i^{th}$ sample is represented as follows

$$(X_i, y_i) = (x_{i1}, x_{i2}, \ldots, x_{id}, y_i)$$

Please note that $y_i$ is still 1-dimensional.

Now, suppose we wish to fit a hyperplane of the following form in $d$ dimensions. Hence for a given $X = (x_1, x_2, .., x_d)$ it should be able to predict a $y$ value using the following equation

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_d x_d$$

In order to fit this hyperplane, we will again have a new definition of $X, Y$ and $\beta$ as follows:

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}; \quad X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1d} \\ 1 & x_{21} & x_{22} & \cdots & x_{2d} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nd} \end{bmatrix}; \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_d \end{bmatrix}$$

And $\beta$ will again be inferred as:

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

Higher degree polynomials etc can also be incorporated directly into the model as before when doing multiple linear regression

# 2 Characterizing a linear regression model

## 2.1 Summarizing discussed forms of linear regression models

With $X, Y$ and $\beta$ representing covariates, observations and parameters/weights, any regression model that can be expressed in the form:

$$Y = X\beta + \epsilon, \quad \epsilon \sim N(0, \sigma^2 I) \tag{5}$$

is a linear model with an estimate of $\beta : \hat{\beta} = (X^T X)^{-1} X^T Y$. Assuming n to be the number of samples, here $Y \in \mathbb{R}^n$, $X \in \mathbb{R}^{nx(p+1)}$, $\beta \in \mathbb{R}^{(p+1)}$, $\sigma^2 \in \mathbb{R}^+$ and $I$ is an n-dimensional identity matrix. Please note that the residual vector: $[e_1, e_2, e_3, e_4]^T$ discussed in the experiments at the beginning of this notebook is just a realization of $\epsilon$ in this general model formulation.

**Hence, all the models discussed above belong to the general family of Linear Regı**
with separate definitions for $X$, $Y$ and $\beta$ as shown in the below mentioned cases.
Please note X matrix formulated for fitting a regression model is generally referred to ε

1. **Case 1**: n 1-dimensional (univariate) samples and fitting a straight line model.
$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}; \quad X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}; \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$$
Here p = 1.
2. **Case 2**: n 1-dimensional (univariate) samples and fitting
$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \ldots + \beta_m x^m$
$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}; \quad X = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^m \\ 1 & x_2 & x_2^2 & \cdots & x_2^m \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^m \end{bmatrix}; \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{bmatrix}$$
Here p = m.
3. **Case 3**: n 1-dimensional (univariate) samples and fitting
$y = \beta_0 + \beta_1 sin(x) + \beta_2 cos(x)$
$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}; \quad X = \begin{bmatrix} 1 & Sin(x_1) & Cos(x_1) \\ 1 & Sin(x_2) & Cos(x_2) \\ \vdots & \vdots & \vdots \\ 1 & Sin(x_n) & Cos(x_n) \end{bmatrix}; \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$$
Here p = 2.
4. **Case 4**: n d-dimensional (multivariate) samples and fitting
$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_d x_d$
$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}; \quad X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1d} \\ 1 & x_{21} & x_{22} & \cdots & x_{2d} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nd} \end{bmatrix}; \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_d \end{bmatrix}$$
Here p = d.

## 2.2 What makes these regression models linear

Hence, for a regression model to be linear, following conditions should be satisfied:

1. $\textbf{Linearity}$: Y and X should be related linearly. One good way to check it is that the model should have the ability to be expressed as a matrix vector product with coeffcients ($\beta$) separable from covariate matrix X.
2. $\textbf{Homoscedasticity}$: The variance of residual is the same for any value of X.
3. $\textbf{Independence}$: Observations are indendent of each other.
4. $\textbf{Normality}$: For any fixed value of X, Y is normally distributed.

$\textbf{Please note}$: For the assumed model definiton in (5) (mentioned below as well) all 4 requirements for a linear regression model are satisfied:

$$Y = X\beta + \epsilon, \quad \epsilon \sim N(0, \sigma^2 I)$$

1. The product $X\beta$ ensures 1 is satisfied (linearity)
2. Error term $\epsilon$ being additive to $X\beta$ with a distribution $N(0, \sigma^2 I)$ gurantees (2), (3) and (4)

## 2.3 Non-linear regression models

All the models that don't satisfy the linear model specific conditions are a kind of a non-linear regression model. For example Neural Network models. These models dont have closed form expressions for parameters/weigths (like $\hat{\beta} = (X^T X)^{-1} X^T Y$ for linear models)

# Exercise

1. Generate data from $f(x) = Sin(x) + log(x)$ for $x \in [0, 10]$
2. Find the weights $\hat{\beta}$ for fitting 4 separate models:

- $y = \beta_0 + \beta_1 Sin(x)$
- $y = \beta_0 + \beta_1 Cos(x),$
- $y = \beta_0 + \beta_1 Sin(x) + \beta_2 log(x)$
- $y = \beta_0 + \beta_1 Cos(x) + \beta_2 log(x)$

3. Predict f(x) for x = 5.7, using the learnt weights $\hat{\beta}$ for each model separately.

In [ ]: