REGRESSION ASSIGNMENT 1

A Proof From Linear Regression

Theorem 1. ***Prove*** *that under Gaussian Assumption Linear Regression Amounts To Least square*

Let us assume that the target variables and the inputs are related via the equation

$$y^{(i)} = \theta^T x^{(i)} + \epsilon^{(i)}$$

where $\epsilon^{(i)}$ is an error term that captures either unmodeled effects , or random noise. Let us further assume that the $\epsilon^{(i)}$ are distributed IID according to a Gaussian distribution (also called a Normal distribution)with mean zero and some variation σ . We can write this assumption as

$$\epsilon^{(i)} \sim N(0, \sigma^2)$$

I.e., the density of ϵ^i is given by

$$p(\epsilon^i) = \frac{1}{\sqrt{2\pi}\sigma} \quad exp\left(-\frac{(\epsilon^i)^2}{2\sigma^2}\right)$$

This implies that

$$p(y^i|X^i;\theta) = \frac{1}{\sqrt{2\pi}\sigma} \quad exp\left(-\frac{(y^{(i)} - \theta^T x^{(i)})^2}{2\sigma^2}\right)$$

The notation $p(y^i|x^i;\theta)$ indicates that this is the distribution of $y^{(i)}$ given $x^{(i)}$ and parameterized by θ. The probability of the data is given by $p(\vec{y}|X;\theta)$. This quantity is typically viewed a function of $\vec{y}$, for a fixed value of θ . When we wish to explicitly view this as a function of θ ,we will instead call it the likelihood function

$$L(\theta) = L(\theta, X, \vec{y} = p(y|\vec{X};\theta)$$

$$= \prod_{i=1}^{m} p(y^i|X^i;\theta) = \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi}\sigma} \quad exp\left(-\frac{(y^{(i)} - \theta^T x^{(i)})^2}{2\sigma^2}\right)$$

Instead of maximizing $L(\theta)$, we can also maximize any strictly increasing function of $L(\theta)$ .In particular , the derivatives will be a bit simpler if we instead maximize the **log likelyhood**

$$l(\theta) = \log L(\theta) = log \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi}\sigma} \quad exp\left(-\frac{(y^{(i)} - \theta^T x^{(i)})^2}{2\sigma^2}\right)$$

$$= \Sigma_{i=1}^{m} log \frac{1}{\sqrt{2\pi}\sigma} \quad exp\left(-\frac{(y^{(i)} - \theta^T x^{(i)})^2}{2\sigma^2}\right)$$

$$mlog \frac{1}{\sqrt{2\pi}\sigma} - \frac{1}{\sigma^2} * \frac{1}{2} \Sigma_{i=1}^{m} \left(y^{(i)} - \theta^T x^{(i)}\right)^2$$

which we recognize to be $j(\theta)$, our original least-squares cost function. |

$$- - - - * * * * * - - - -$$