Lecture 12 (Newton's Method)

1 Newton's Method for Optimisation

1. Uses 2^{nd} order information of the cost functions to make *faster* progress
2. Uses the intersection of tangent with "x" axis to approach towards the zeros of function
3. $\theta^{(t+1)} = \theta^t - \frac{h\left(\theta^t\right)}{h'\left(\theta^t\right)}$
4. Now, replace $h(\theta) = \nabla_\theta J(\theta)$
5. For multi-variable, the equation changes to

$$\theta^{(t+1)} = \theta^t - \left(H^{-1}\nabla_\theta J(\theta)\right)\Big|_{\theta^t}$$

2 Locally Weighted Linear Regression

1. These are non-parametric and lazy methods
2. Learns multiple linear functions instead of finding polynomial solution for non-linear training data
3. It doesn't actually "learn" in advance but performs computation when input is given
4. $J^x(\theta) = \frac{1}{2m}\sum_{i=1}^{m}\left(w_i\left(y_i - h_\theta(x_i)\right)^2\right)$, where w_i is inversely proportional to distance of input data to each x_i in training data
5. $w_i = \exp\left(\frac{-(x-x_i)^2}{2\tau^2}\right)$ is a good choice
6. For multi-variate case, $w_i = \exp\left(\frac{-(x-x_i)^T\Sigma^{-1}(x-x_i)}{2}\right)$