Lecture 10 (Analytical Solution)

1 Plot of $J(\theta)$ for SGD

The graph has extra *zig-zag* points between each epoch (once entire batch is consumed). The graph has *periodic* convergence. So the algorithm is stopped only at number of iterations which are multiples of m/r.

2 Analytical Solution for Least Square Regression

Design matrix $X \in \mathbb{R}^{m\times(n+1)}$ is such that each row X_i is x_i^T. $Y \in \mathbb{R}^m$ is vector of of y_i. $\Theta \in \mathbb{R}^{n+1}$ is vector of θ_i. Now consider,

$$X\Theta - Y$$

This is the *difference* part of the error. Thus, $J(\theta)$ equals

$$\frac{1}{2m}(X\Theta - Y)^T(X\Theta - Y)$$

On solving the equation, $\nabla_\theta J(\theta) = 0$ (and simplifying the equation),

$$\frac{1}{2m}\nabla_\theta \left(\Theta^T X^T X\Theta - 2\Theta^T X^T Y + Y^T Y\right) = 0$$

$$\implies \frac{1}{m}X^T\left(X\Theta - Y\right) = 0$$
$$\implies \Theta = (X^T X)^{-1}X^T Y$$