Multiple Linear Regression

$$y = X\theta$$

$$X^T y = X^T X \theta$$

$$\theta = (X^T X)^{-1} X^T y$$

Ridge Regression

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} \left(y_i - \theta^T x_i\right)^2 + \underbrace{\frac{\lambda}{2m} \sum_{j=1}^{n} \theta_j^2}_{\text{Penalty}}$$

ignore $1/2m$

$$J(\theta) = \|y - X\theta\|^2 + \lambda \|\theta\|^2$$

$$= (y - X\theta)^T (y - X\theta) + \lambda \theta^T \theta$$

$$= y^T y - y^T X\theta - X^T \theta^T y + X^T X \theta \theta^T + \lambda \theta^T \theta$$

$$\frac{\partial(J(\theta))}{\partial\theta} = -2X^T y + 2\theta\left(X^T X + \lambda\right)$$

To minimise $J(\theta)$

$$\frac{\partial}{\partial\theta} = 0$$

Gradient Descent

$$2X^T y = 2\theta\left(X^T X + \lambda\right)$$

$$\theta = \left(X^T X + \lambda\right)^{-1} X^T y$$

- Lasso

$$J(\theta) = \frac{1}{2m}\left(Y - \theta^T X\right)^T \left(y - \theta^T X\right) + \frac{\lambda}{m}|\theta| \quad \text{Penalty}$$

as we can't differentiate $|\theta|$ at 0 so this is not a closed form & we use coordinate descent,

Coordinate Descent

Minimise multivariable function by updating one parameter at a time, holding all other parameters fixed.