Alvin Karanja - u2156001

Question 8

- Data : z

- Data is partitioned by predictor variables : x, y

- Resulting in the following model :

$$\hat{z} = ax^2 + by^2 + cxy + d$$

- Model follows regression paradigm such that :

$$\begin{bmatrix} \hat{z}_i \\ \vdots \\ \hat{z}_n \end{bmatrix} = \begin{bmatrix} x_i^2 + y_i^2 + xy_i + 1 \\ \vdots \\ x_n^2 + y_n^2 + xy_n + 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}$$

- We can consider our model as a function parameterised by : a, b, c, d

$$\therefore \quad \hat{z} = \hat{z}(a, b, c, d)$$

- We can consider the cost of our model the difference between actual values and predicted values

$$\text{Cost} = z - \hat{z}$$

→ To handle +ve and -ve error values, we can take the sum of the squared errors

$$\varepsilon = \sum (z - \hat{z})^2$$

$\therefore$ cost function ε is sum of squared errors.

- Substituting for our model function:

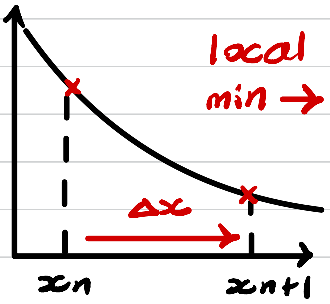
$* \ \hat{z} = ax^2 + by^2 + cxy + d$

$$\mathcal{E} = \sum \left(z - (ax^2 + by^2 + cxy + d) \right)^2$$

→ Thus we can consider our cost function parameterised by: a, b, c, d

Aside – Gradient Descent

- Consider two points x_n & x_{n+1}:

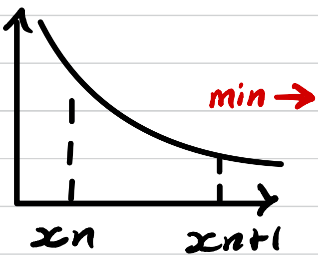


In order to approach the local minimum, we must step:

$$x_n \longrightarrow x_{n+1}$$

This can be done by moving Δx given by $\frac{d}{dx}$.

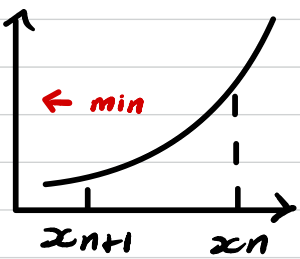
case a.) Consider the following:



The value of $\frac{d}{dx_n}$ is $-$ve, thus:

$x_n - \left(-\frac{d}{dx_n} \right)$: Move towards min by $+\Delta x$

case b.) Alternitively consider:



The value of $\frac{d}{dx_n}$ is $+$ve, thus:

$x_n - \left(\frac{d}{dx_n} \right)$: Move towards min by $-\Delta x$

- At the minimum point:

$$\frac{d}{dx} = 0 \quad \therefore \quad x_{n+1} = x_n - 0$$

$$\therefore \quad x_{n+1} = x_n$$

- Thus in order to minimise a function, can compute:

Repeat:

$$x_{n+1} = x_n - \frac{d}{dx}$$

Until: $x_{n+1} = x_n$ (convergence)
or N iterations

- Additionally, steps in Δx can be in fractions η

* Thus gradient descent can by accomplished by following:

$$x_{n+1} = x_n - \eta \frac{d}{dx}$$

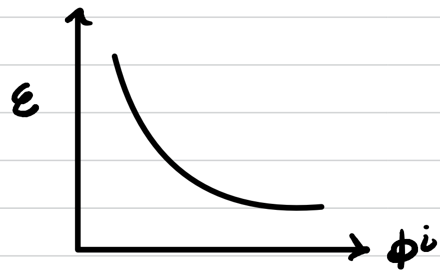
- Consider cost function:

$$\varepsilon = \sum (z - (ax^2 + by^2 + cxy + d))^2$$

- Wish to minimise cost by minimising parameters a, b, c, d

- We can consider each dimension of the cost function such that:

$\phi^i \in \{a, b, c, d\}$



- Thus for each dimension we can approach the minimum via the gradient descent algorithm:

$$\phi^i_{k+1} = \phi^i_k - \eta \frac{\partial \varepsilon}{\partial \phi^i} \qquad \phi^i \in \{a, b, c, d\}$$

→ we can thus consider the following gradients:

$$\frac{\partial \varepsilon}{\partial a} = -2x^2 \sum \left(z - (ax^2 + by^2 + cxy + d) \right)$$

$$\frac{\partial \varepsilon}{\partial b} = -2y^2 \sum \left(z - (ax^2 + by^2 + cxy + d) \right)$$

$$\frac{\partial \varepsilon}{\partial c} = -2xy \sum \left(z - (ax^2 + by^2 + cxy + d) \right)$$

$$\frac{\partial \varepsilon}{\partial d} = -2 \sum \left(z - (ax^2 + by^2 + cxy + d) \right)$$

- We can thus apply the gradients in the descent loop for each parameter

$$a_{k+1} = a_k - \eta \frac{\partial \varepsilon}{\partial a_k}$$

$$b_{k+1} = b_k - \eta \frac{\partial \varepsilon}{\partial b_k}$$

$$c_{k+1} = c_k - \eta \frac{\partial \varepsilon}{\partial c_k}$$

$$d_{k+1} = d_k - \eta \frac{\partial \varepsilon}{\partial d_k} \qquad \#$$

→ Repeat till convergence or for N iterations.