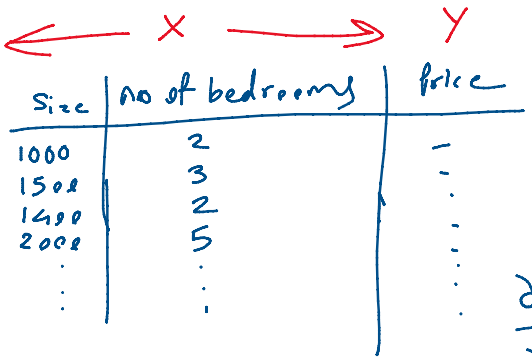
Feature Scaling ML GURU



$$\hat{y}_i = w_1 x_{i1} + w_2 x_{i2} + b$$

From Previous video,

$$\frac{\partial L}{\partial w_j} = \frac{1}{m}\left[(\hat{y}_1 - y_1)x_{1j} + (\hat{y}_2 - y_2)x_{2j} + \cdots (\hat{y}_m - y_m)x_{mj}\right]$$

$$\frac{\partial L}{\partial w_j} \propto (\hat{y}_i - y_i)x_{ij}$$

eg)

$$\frac{\partial L}{\partial w_1} \propto (\hat{y}_i - y_i)x_{i1}$$

$$\frac{\partial L}{\partial w_2} \propto (\hat{y}_i - y_i)x_{i2}$$

(x_{i1} (1000's))

$x_{i1} \Rightarrow$ (size)

$x_{i2} \Rightarrow$ (no. of bedrooms)

$$\frac{\partial L}{\partial b} = (\hat{y}_i - y_i)(1)$$

$$\therefore \left|\frac{\partial L}{\partial w_1}\right| >> \left|\frac{\partial L}{\partial w_2}\right|$$

update rule:

$$w_1 = w_1 - lr\left(\frac{\partial L}{\partial w_1}\right)$$

$$w_2 = w_2 - lr\left(\frac{\partial L}{\partial w_2}\right)$$

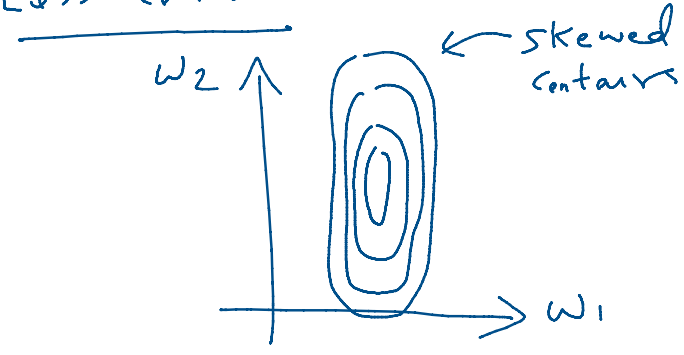
$$\Delta w_1 \propto \frac{\partial L}{\partial w_1}$$

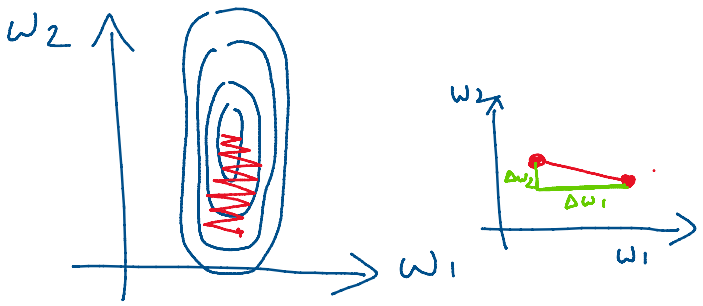
$$\Delta w_2 \propto \frac{\partial L}{\partial w_2}$$

Hence,

$$\Delta w_1 >> \Delta w_2$$

Loss contour:





Hence, All the features must be in same range

Feature Scaling

① Feature normalization (MinMax scaler)

→ It scales all the features (columns) to [0,1] range

eg)

no. of bedrooms
1
3
5
2

$X_{min} = 1$

$X_{max} = 5$

$$X = \frac{X - X_{min}}{X_{max} - X_{min}}$$

eg) $X = 5$

$$\therefore X_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}} = \frac{5-1}{5-1} = 1$$

$X = 3$

$$X_{scaled} = \frac{3-1}{5-1} = \frac{2}{4} = 0.5$$

② Feature Standardisation

→ scales all features to 0 mean and unit variance

$$X = \frac{X - \mu}{s}$$

μ = mean
s = standard deviation