

Squared Errors

$$E = \frac{1}{2} \sum_{\mu} \sum_j [y_j^{\mu} - \hat{y}_j^{\mu}]^2$$

(SSE)
sum of squared errors

Annotations:
- μ : data points
- j : output units
- y_j^{μ} : true value
- \hat{y}_j^{μ} : prediction
- prediction depends on w_{ij} 's

Gradient is another word for slope or rate of change.

Gradient descent can get into "local minima".

One way to solve is "momentum".

error term: $\delta = (y - \hat{y}) f'(h)$

\hookrightarrow activation function

$$w_i = w_i + \eta \delta x_i$$

$$h = \sum w_i x_i$$

Mean Squared Error (MSE)

$$E = \frac{1}{2m} \sum_{\mu} (y^{\mu} - \hat{y}^{\mu})^2$$