Stochastic gradient descent

loop {

for $i = 1$ to $i = m$ {

$$\theta_j := \theta_j - \alpha(h_\theta(x^{(i)}) - y^{(i)})\, x_j^{(i)}$$

(for $j = 0 \to n$)

}

}

Useful for large datasets (i.e large value of m)

Matrix derivatives

$f: \mathbb{R}^{m \times n} \mapsto \mathbb{R}$

$$A = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & & & A_{2n} \\ \vdots & & & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{bmatrix}_{m \times n}$$

gradient of $f(A)$ w.r.t A

$$\nabla_A f(A) = \begin{bmatrix} \frac{\partial f}{\partial A_{11}} & \frac{\partial f}{\partial A_{12}} & \cdots & \frac{\partial f}{\partial A_{1n}} \\ \frac{\partial f}{\partial A_{21}} & \frac{\partial f}{\partial A_{22}} & \cdots & \frac{\partial f}{\partial A_{2n}} \\ \vdots & & & \vdots \\ \frac{\partial f}{\partial A_{m1}} & \frac{\partial f}{\partial A_{m2}} & & \frac{\partial f}{\partial A_{mn}} \end{bmatrix}_{m \times n}$$

derivative of f w.r.t A

$\to (i,j)^{th}$ element $\frac{\partial f}{\partial A_{ij}}$

If A is a 2×2 matrix $\Rightarrow A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$

and func^n $f: \mathbb{R}^{2\times 2} \to \mathbb{R}$

$$f(A) = \frac{3}{2} A_{11} + 5A_{12}^2 + A_{21} A_{22}$$

$$\nabla_A f(A) = \begin{bmatrix} \frac{3}{2} & 10A_{12} \\ A_{22} & A_{21} \end{bmatrix}$$

Trace of a matrix

$A \to n \times n$ matrix ~~(only~~

$$\operatorname{tr} A = \sum_{i=0}^{n} A_{ii} = A_{11} + A_{22} + \cdots + A_{nn}$$ (only for square matrix)