H.W-5

RAKSHITHA PANDURANGA
UCC ID - 7890-1614-34
rpandura@uncc.edu

2a) $t_n(N) = \frac{1}{N-1+1} \qquad 1 \le n \le N$

$$E\{\Delta \underline{w}(i)\} = E(-\eta \nabla_{\underline{w}} J_n(\underline{w}))$$

$$= \sum_{1}^{N} -\eta \nabla_{\underline{w}_i} J_n(\underline{w}_i) \times \frac{1}{N}$$

$$\boxed{= \frac{-\eta}{N-1} \sum \nabla_{w_i} J_n(\underline{w})}$$

b). $E\left\{\sum_{i=0}^{N-1} \Delta \underline{w}(i)\right\}$

linearity $\sum_{i=0}^{N-1} E(\Delta \underline{w}(i))$

$$= \sum_{i=0}^{N-1} -\frac{\eta}{N} \sum_{n=1}^{N} \nabla_{\underline{w}} J_n(\underline{w})$$

$$\boxed{E\left\{\sum_{i=0}^{N-1} \Delta \underline{w}(i)\right\} = -\eta \sum_{n=1}^{N} \nabla_{\underline{w}} J_n(\underline{w}) = -\eta \nabla_{\underline{w}} J(\underline{w})}$$

c) Batch gradient descent

$$\underline{w}(i+1) = \underline{w}(i) - \eta(i) \nabla_{\underline{w}} J(\underline{w})$$

$$\Delta \underline{w}(i) = -\eta \nabla_{\underline{w}} J(\underline{w})$$

$$E[\Delta \underline{w}(i)] = -\eta E[\nabla_{\underline{w}} J(\underline{w})]$$

$$\therefore -\eta \nabla_{\underline{w}} J(\underline{w})$$

$$E\left[\sum_{i=0}^{N-1} \Delta \underline{w}(i)\right] = N E[\Delta \underline{w}(i)]$$

$$= -\eta N \nabla_{\underline{w}} J(\underline{w})$$

Batch gradient descent

→ Average of all weight updates is N times that of stochastic gradient descent Variant 2

Stochastic gradient descent

→ Average of all weight updates is independent of total number of points N, and depends on total no. of points in batch gradient descent.

Additionally,

$$E\left\{\sum_{i=0}^{N-1} \Delta \underline{w}(i)\right\} = \Delta \underline{w}(i)$$

↓ Stochastic gradient descent

↓ Batch gradient descent