MLFoundation HW4

r07922100 楊力權

1.

2.

gradient descent : $w_{t+1} \leftarrow w_t - \eta \nabla E_{in}$

augmented error gradient descent : $w_{t+1} \leftarrow w_t - \eta \nabla E_{aug}$

$\because E_{aug} = E_{in} + \frac{\lambda}{N} w^T w \quad \therefore \nabla E_{aug} = \nabla E_{in} + \frac{\partial}{\partial w} \frac{\lambda}{N} w^T w = \nabla E_{in} + \frac{2\lambda}{N} w$

因此得$w_{t+1} \leftarrow w_t - \eta \nabla E_{aug} = w_t - \eta(\nabla E_{in} + \frac{2\lambda}{N} w_t) = (1 - \frac{2\eta\lambda}{N}) w_t - \eta \nabla E_{in}(w_t)$

3.

$(-1,0), (\rho,1), (1,0)$三點一次取出一點當作LOO則會得三條線：

$(-1,0), (1,0) : h_1(x) = 0$

$(-1,0), (\rho,1) : h_1(x) = \frac{1}{\rho+1} x + \frac{1}{\rho+1}$

$(1,0), (\rho,1) : h_1(x) = \frac{1}{\rho-1} x + \frac{1}{1-\rho}$

可得

$$E_{loo} = \frac{1}{3}((\frac{-1}{\rho-1} + \frac{1}{1-\rho})^2 + (\frac{1}{\rho+1} + \frac{1}{\rho+1})^2 + 1) = \frac{1}{3}(4(\frac{1}{\rho-1})^2 + 4(\frac{1}{\rho+1})^2 + 1) = \frac{1}{3}(4\frac{2\rho^2+2}{(\rho^2-1)^2} + 1)$$

$$= \frac{1}{3}(\frac{\rho^4 + 6\rho^2 + 9}{\rho^4 - 2\rho^2 + 1})$$

4.

SGD on original & virtual samples

$X = [x_1, x_2, \cdots]^T, Y$ 為original samples

$\tilde{X} = \sqrt{\lambda} I = [[\sqrt{\lambda},0,\cdots],[0,\sqrt{\lambda},0,\cdots],\cdots]$, $\tilde{Y} = [0,0,\cdots]$為virtual samples

for i in iterations

random pick a sample from $X,\ \tilde{X}$ with its label in $Y,\ \tilde{Y}$

calculate the gradient ∇E_{in} with Squared Error function

update w with rule $w_{t+1} \leftarrow w_t - \eta \nabla E_{in}(w_t)$

Augmented gradient descent

$$w_{t+1} \leftarrow (1 - \frac{2\eta\lambda}{N})w_t - \eta \nabla E_{in}(w_t)$$

使用N+K筆資料是透過加入的virtual samples調整梯度來控制$\|w\|^2$的成長

使用Augmented的GD是透過每個iteration都削減原本的weight來控制$\|w\|^2$

5.

error = $|sin(ax) - wx| = \sqrt{(\sin(ax) - wx)^2} = \sqrt{\sin^2(ax) + (wx)^2 - 2wx\sin(ax)}$

$if\ \ w \neq 0\ \ then\ \ \lim_{x\to\infty} (wx)^2 - 2wx\sin(ax) = \infty\ \ and\ \ then$ error $\rightarrow \infty$

$if\ \ w = 0\ \ then\ \ \lim_{x\to\infty} (0x)^2 - 2\times 0x\sin(ax) = 0\ \ and\ \ then$ error $= |\sin(ax)|$

而deterministic noise是f與 best hypothesis $h*$ 間的誤差，得

deterministic noise = $|sin(ax)|$