Problem 1 (5+5+10)

Consider a function:

$$E(w_0w_1) = e^{w_0} + e^{w_1} + e^{w_0} - 3w_0 + w_1^2 - 2w_1 + 5 \quad \text{—(1)}$$

a) Equivalent $E(w_0w_1, \lambda w_0w_1)$ by $\lambda$ and $\lambda^2$. Since $E$ is the RSS cost function, equivalent around $(w_0, w_1) = (0,0)$. Suppose $E(w_0, w_1) = w_0, w_1 = w_0, w_1 = w_1$

and we evaluate $E$ at $(\lambda w_0, \lambda w_1)$

substituting $(\lambda w_0, \lambda w_1)$ and $\lambda^2 E(w_0, w_1)$ into

$$E(\lambda w_0, \lambda w_1) = e^{\lambda w_0} + e^{\lambda w_1} + e^{\lambda w_0} - 3\lambda w_0 + \lambda^2 w_1^2 - 2\lambda w_1 + 5$$

$$\lambda^2 E(w_0, w_1) = \lambda^2 e^{w_0} + \lambda^2 e^{w_1} + \lambda^2 e^{w_0} - 3\lambda^2 w_0 + \lambda^2 w_1^2 - 2\lambda^2 w_1 + 5\lambda^2$$

$$\Rightarrow \underbrace{e^{\lambda w_0} + e^{\lambda w_1} + e^{\lambda w_0} - 3\lambda w_0 + \lambda^2 w_1^2 - 2\lambda w_1 + 5}_{\text{not equivalent}} \neq \lambda^2 e^{w_0} + \lambda^2 e^{w_1} + \lambda^2 e^{w_0} - 3\lambda^2 w_0 + \lambda^2 w_1^2 - 2\lambda^2 w_1 + 5\lambda^2$$

$$\Rightarrow \underbrace{\lambda e^{w_0} + \lambda e^{w_1} + \lambda e^{w_0} - 3\lambda w_0 + \lambda w_1^2 - 2\lambda w_1 + 5\lambda}_{\text{not equivalent}} \neq \lambda^2 e^{w_0} + \lambda^2 e^{w_1} + \lambda^2 e^{w_0} - 3\lambda^2 w_0 + \lambda^2 w_1^2 - 2\lambda^2 w_1 + 5\lambda^2$$

$$E(\lambda w_0, \lambda w_1) \neq \lambda^2 E(w_0, w_1)$$

$$\Rightarrow \underbrace{\lambda e^{w_0} + \lambda e^{w_1} + \lambda e^{w_0} - 3\lambda w_0 + \lambda w_1^2 - 2\lambda w_1 + 5\lambda}_{\text{not equivalent}} \neq \lambda^2 e^{w_0} + \lambda^2 e^{w_1} + \lambda^2 e^{w_0} - 3\lambda^2 w_0 + \lambda^2 w_1^2 - 2\lambda^2 w_1 + 5\lambda^2$$

$$\Rightarrow \underbrace{\lambda e^{w_0} + \lambda e^{w_1} + \lambda e^{w_0} - 3\lambda w_0 + \lambda w_1^2 - 2\lambda w_1 + 5\lambda}_{\text{not equivalent}} \neq \lambda^2 e^{w_0} + \lambda^2 e^{w_1} + \lambda^2 e^{w_0} - 3\lambda^2 w_0 + \lambda^2 w_1^2 - 2\lambda^2 w_1 + 5\lambda^2$$

$$\Rightarrow \underbrace{\lambda e^{w_0} + \lambda e^{w_1} + \lambda e^{w_0} - 3\lambda w_0 + \lambda w_1^2 - 2\lambda w_1 + 5\lambda}_{\text{not equivalent}} \neq \lambda^2 e^{w_0} + \lambda^2 e^{w_1} + \lambda^2 e^{w_0} - 3\lambda^2 w_0 + \lambda^2 w_1^2 - 2\lambda^2 w_1 + 5\lambda^2$$

$$E(\lambda w_0, \lambda w_1) \neq \lambda^2 E(w_0, w_1)$$

b) Minimize $\hat{E}_1$ from all possible $(w_0, w_1)$ such that $E(w_0, w_1) = 0$. So this implies we need find the optimal where under $\{\hat{w}_0^*\}$ is equal to the values $E(w_0, w_1)$ which is called regular optimal direction.

Compute the optimal $E(w_0, w_1)$ and the matching $E(w_0, w_1)$

$$\hat{E}_1 = -2.36 + 0.64 + 3$$

$$E(w_0, w_1) = 5.5 \quad \text{so this can be possible where}$$

$$\text{Subject to} \begin{cases} 3w_0 + 2w_1 \\ 2w_1 - 2 = 0 \end{cases}$$

$$\text{Find } \{w_0, w_1\} = \begin{bmatrix} \frac{3}{2}(\ln 2 - 1 + \sqrt{1 - \ln 2}) \\ 1 \end{bmatrix} = \begin{bmatrix} 0.5 \\ 1 \end{bmatrix}$$

$$\hat{E}_1^* = -2.(0.5)^{0.5} - 2(1)^2(0.5) + 3$$

$$E = -1.414 - 1.5 + 3 = 0$$

$$\text{minimize } \hat{E}_1^* \text{ and take derivative wrt } \theta \text{ and equate to } 0.$$

$$\hat{E}_1^* = 1.414 - 1.5 \cdot 0.5 = 0$$

$$0.414 = 0.75$$

using $\sin^2\theta + \cos^2\theta = 1$ we find $\cos\theta$ and $\sin\theta$

$$\frac{1}{2}\sin^2\theta + \cos^2\theta = 1$$

$$\frac{1}{2}\sin^2\theta = 1$$

$$\sin^2\theta = 2$$

$$\cos^2\theta = 1$$

$$\sin\theta = \left(\frac{1}{2}\right)^{0.5} = 1$$

$$\cos\theta = \frac{1}{2}\cos^2\theta = 1$$

$$\frac{1}{2}\cos^2\theta = 1$$

$$\cos^2\theta = \frac{2}{1}$$

$$\cos\theta = \sqrt{2}$$

optimal for $\sin\theta$ and $\cos\theta$

$$\sin\theta = 0.5 + 0.5 = \frac{1}{2}$$

$$\cos\theta = 0.5 + 0.5 = \frac{1}{2}$$

computing for matching $E(w_0, w_1)$

Taylor series expansion

$$\nabla E(w_0) = \begin{bmatrix} \frac{\partial E_{w_0}}{\partial w_0} \\ \frac{\partial E_{w_1}}{\partial w_1} \end{bmatrix}$$

$$= \begin{bmatrix} e^{w_0} + e^{w_0} - 3 \\ 2e^{w_1} + 2e^{w_0} - 3w_0 + 2w_1 - 2 \end{bmatrix}$$

at $(w_0, w_1) = (0,0)$

$$\nabla E(w_0) = \begin{bmatrix} -1 \\ -2 \end{bmatrix}$$

c) Approximate $E(w_0, w_1)$ by $\hat{E}_2(w_0, w_1)$ where $\hat{E}_2$ is the second order Taylor series expansion of $E$ around $(w_0, w_1) = (0,0)$

Suppose $\hat{E}_2 = E_0 + (w_0 - 0)^T \frac{\partial E}{\partial w_0} + \frac{1}{2}(w_0 - 0)^T H(w_0 - 0) + \frac{1}{6} w_0 w_1 + \frac{1}{6} w_0 w_1 + \frac{1}{6}$

find values of $E_0, \frac{\partial E}{\partial w_0}, w_0, w_1, H_{w_0}, H_{w_1}$

$$E(w_0, w_1) = e^{w_0} + e^{w_1} + e^{w_0} - 3w_0 + w_1^2 - 2w_1 + 5$$

$$= (w_0 + w_1)^3 - 3(w_0 + w_1)(w_0 + w_1) + 3(w_0 + w_1) - 2(w_0 + w_1) + 5$$

Using second order Taylor expansion for $E(w_0, w_1)$

$$\rightarrow \frac{\partial E(w_0)}{\partial w_0} = \underbrace{(w_0 + w_1)^2 - 3(w_0 + w_1)}_{\text{optimal}} \quad \text{—(2)}$$

$$\rightarrow \frac{\partial E(w_0)}{\partial w_1} = \underbrace{2(w_0 + w_1) - 3(w_0 + w_1)}_{\text{optimal}} \quad \text{—(3)}$$

$$\rightarrow \frac{\partial E(w_0)}{\partial w_0 \partial w_1} = \underbrace{(w_0 + w_1)^2 - 3(w_0 + w_1)}_{\text{optimal}} \quad \text{—(4)}$$

$$E(w_0, w_1) = (w_0 + w_1)^3 - 3(w_0 + w_1)(w_0 + w_1) + 3(w_0 + w_1) - 2(w_0 + w_1) + 5$$

$$\text{at } (w_0, w_1) = (0,0) \quad \rightarrow \nabla E(w_0, w_1) = \begin{bmatrix} -1 \\ -2 \end{bmatrix}$$

$$= \left[(-1 + 2w_1 + 2w_0)^2 + \left(-2 + 2w_1 + 2w_0\right)^2 + \left(-1 + 2w_1 + 2w_0\right)^2\right]$$

$$= 8w_0^2 + 4w_0^2 - 12w_0 + 2w_0 - 3w_0 + 5$$

$$\boxed{w_0 = 1, w_1 = 0, E(w_0, w_1) = 5}$$

d) Minimize $\hat{E}_2$ from all possible $(w_0, w_1)$. Use the find that $\nabla E(w_0)_{w_0}$ to see objects to prove that the optimal where under

$$\begin{bmatrix} w_0^* \\ w_1^* \end{bmatrix} = -(\nabla^2 E(w_0))^{-1} \nabla E(w_0)$$

which is the Newton Iteration.

finding to minimize $\hat{E}_1$ → take derivative and equate to 0

$$\frac{\partial \hat{E}_2}{\partial w_0} = \frac{1}{2}(3\sqrt{w_0^2 + w_1^2} - 2w_0 - 2w_1 - 3\sqrt{w_0^2 + w_1^2})^2$$

$$= 3w_0 - 2w_0 - 2w_1 = 0$$

$$\frac{\partial \hat{E}_2}{\partial w_1} = \frac{1}{2}(3\sqrt{w_0^2 + w_1^2} - 2w_0 - 2w_1 - 3\sqrt{w_0^2 + w_1^2})^2$$

$$= -2w_0 + 12w_1 - 6 = 0$$

$$\begin{bmatrix} \frac{\partial^2 \hat{E}_2}{\partial w_0^2} & \frac{\partial^2 \hat{E}_2}{\partial w_0 \partial w_1} \\ \frac{\partial^2 \hat{E}_2}{\partial w_1 \partial w_0} & \frac{\partial^2 \hat{E}_2}{\partial w_1^2} \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$$

$$\begin{bmatrix} -2 & -2 \\ -2 & 12 \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$$

$$\rightarrow -\nabla E(w_0)$$

$$\rightarrow \begin{bmatrix} w_0 \\ w_1 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \end{bmatrix} \begin{bmatrix} -2 & -2 \\ -2 & 12 \end{bmatrix}$$

$$\begin{bmatrix} w_0 \\ w_1 \end{bmatrix} = \begin{bmatrix} \frac{3}{2} \\ \frac{2}{12} \end{bmatrix}$$

Computing $\nabla^2 E(w_0)$ at $(w_0)$

Taylor series expansion $\nabla E(w_0)$ at $(w_0, 0)$ is

$$\nabla E(w_0) + \nabla E(0,0) + \nabla^2 E(w_0) \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}$$

we minimized this previously to get

$$\text{sufficient}$$

$$\nabla E(w_0) + \nabla^2 E(w_0) \begin{bmatrix} w_0 \\ w_1 \end{bmatrix} = 0 \quad \text{—(1)}$$

Now finding the Hessian matrix which is positive semidefinite

$$\nabla^2 E(w_0) = \begin{bmatrix} \frac{\partial^2 E(w_0)}{\partial w_0^2} & \frac{\partial^2 E(w_0)}{\partial w_0 \partial w_1} \\ \frac{\partial^2 E(w_0)}{\partial w_1 \partial w_0} & \frac{\partial^2 E(w_0)}{\partial w_1^2} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\partial^2 E(w_0)}{\partial w_0^2} & \frac{\partial^2 E(w_0)}{\partial w_0 \partial w_1} \\ \frac{\partial^2 E(w_0)}{\partial w_1 \partial w_0} & \frac{\partial^2 E(w_0)}{\partial w_1^2} \end{bmatrix}$$

$$= \begin{bmatrix} 2(-3w_0 + 3) & 2(-3w_0 + 3) \\ 2(-3w_0 + 3) & 2(-3w_0 + 3) + 2 \end{bmatrix}$$

$$= \begin{bmatrix} 3 & -2 \\ -2 & 12 \end{bmatrix}$$

from (1) we can say that if $(w_0) = (w_0)$

$$\nabla E(w_0) + \nabla^2 E(w_0) \begin{bmatrix} w_0 \\ w_1 \end{bmatrix} = 0$$

$$\Rightarrow \begin{bmatrix} w_0 \\ w_1 \end{bmatrix} = -(\nabla^2 E(w_0))^{-1} \nabla E(w_0) \quad \text{from part (b)}$$

$$= \begin{bmatrix} 3 & -2 \\ -2 & 12 \end{bmatrix}^{-1} \begin{bmatrix} -1 \\ -2 \end{bmatrix}$$

Matrix multiplication in Wolfram alpha

$$= \begin{bmatrix} \frac{29}{34} \\ \frac{11}{34} \end{bmatrix}$$

Problem 5 (10)

Deterministic Noise depends on $H$, as some models appropriate better than others.

a) Minimum $H$ is fixed and we increase the complexity of $H$ with deterministic noise in general go up or down?

It may or higher or lower depending on $H$ and $D$.

Ans: From this question it is understood that the complexity of hypothesis $H$ is fixed.

i) It can be fixed but initially $H$ is at less complex than $f$ and we keep increasing the complexity of $H$ the deterministic noise goes down initially but as $H$ becomes more complex than $f$ the deterministic noise goes up and keeps increasing.

ii) It can be fixed but initially $H$ is more complex than $f$ and we keep increasing the complexity of $H$. The deterministic noise goes down initially but as $H$ becomes more complex than $f$ the deterministic noise goes up and keeps increasing.

b) Assume that $H$ is fixed and we decrease the complexity of $H$. Will deterministic noise in general go up or down? It depends on $H$ and a higher or lower starting to overfit.

Ans: i) If we approach $H$ as less complex than $f$ and keep decreasing the complexity of $H$. The deterministic noise goes down initially but as $H$ becomes more complex than $f$ the deterministic noise goes up and keeps increasing. Will $H$ can properly fit $f$ and then $H$ becomes more complex compared to $f$ the deterministic noise increases. Starting to a higher starting to overfit.

Problem 6

For a linear regression problem, find the best set of weights given by the least squares. Given the error that the weights is called the corresponding solution.

other ways obtained. $w_0$ so you form all weights, the state is called the corresponding solution.

Regularization is done basically when there exists a constraint which getting too crowded, we want the $w$ constraint subject

$$\underbrace{w^T w}_{\text{weight}} \leq \underbrace{C}_{\text{constraint}}$$

this is referred to as soft-order constraint as it gives the weights by small or close without changing the order of the optimal.

The goal with defining the optimal weight sign with some defined weights (high order) this will determine the amount of regularization that need to be done. Hence weight.

$$W_{reg} \in W_{\infty} \Leftrightarrow \text{optimal regularized weight}$$

$W_{\infty}$ is calculated using soft regularization parameter in a way that find the weight error is minimized within the defined constraint.

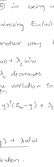
The Insample error for a transformed $\Phi$ is given by.

$$E_{in}(w) = \frac{1}{N}(\Phi w - y)^T(\Phi w - y)$$

minimize

Assume $(x_1, y_1)$ and $(x_2, y_2)$ are not unique that define the weight $w_{reg}$.

If $0$ is small then and $w_{reg}$




Here the radius of $w_{reg}$ is 'C' cannot be increased to find $w_{reg}$ which gives smallest error.

By plotting the level sets we see which does not actually satisfy the constraint. Hence we find the best weight given and which the minimum $w$ to a designated point.

The designed will be a specific solution where dependent will be the weight that is found to specify under a constraint.

The business what is optimal $w$ should be $\Phi$ the $w$. The $w$ do at some along the contour and decrease the

when looks at $w$ for $E_{in}$, it is also $\Phi$ for overall.

$$\Rightarrow \nabla E_{in} \propto -w$$

If $E_{in}$ must be the optimal, then for some the gradient $\nabla E_{in}$

$$\Rightarrow \nabla E_{in}(w_{reg}) = -2\lambda w_{reg}$$

$$\Rightarrow \nabla E_{in}(w_{reg}) \propto 2\lambda w_{reg}$$

$$\text{With gradient convergence}$$

$$\Rightarrow \underbrace{\nabla E_{in}(w)}_{\text{grad}} + \underbrace{2\lambda w(w)}_{\text{grad}} = 0$$

This indicates correct result $\Phi$ is using unconditionally optimal.

→ Now plugging $\lambda$ in and minimizing $E_{in}(w)$ subject to:

This can also be done in another way by plugging $\lambda$ and minimizing

$$E_{aug}(w) = E_{in}(w) + \lambda_i w^T w$$

Let $\lambda = \frac{\lambda}{N}$

$$E_{aug} = \frac{(\Phi w - y)^T(\Phi w - y) + \lambda w^T w}{N}$$

defined using general decision.

Now unconditionally minimizing $E_{aug}(w)$ using gradient

$$\nabla E_{aug}(w) = \frac{2}{N}(\Phi^T \Phi w - \Phi^T y) + 2\lambda w = 0$$

$$= 2(\Phi^T \Phi + \lambda I)w - 2\Phi^T y = 0$$

$$(\Phi^T \Phi + \lambda I)w = \Phi^T y$$

$$\boxed{w_{reg} = (\Phi^T \Phi + \lambda I)^{-1} \Phi^T y}$$

This way we can find regularized weights.

Problem complete