CS5590: Foundations of Machine Learning

Assignment 1

Authors

Name	Roll number
Raj Patil	CS18BTECH11039
Karan Bhukar	CS18BTECH11021

Clarification: using subscripts when indexing into a vector and superscripts when refering to a particular element in a collection

---

Question 1

given the linear model:

$$\hat{Y}^n = w_0 + \sum_{i=1}^{D} w_i X_i^n$$

with loss function as:

$$E_d = \frac{1}{2}||\hat{Y} - Y||_2^2$$

> where $\hat{Y}$ is the prediction vector and Y represents the actuals; their length being N.

Now noise ($\epsilon_i \sim \mathcal{N}(0, \sigma^2)$) is added to each input (x_i) independently. Let ϵ represent the d-random vector. Then each noisy $x' \triangleq x + \epsilon$ where $\epsilon \triangleq [\epsilon_1, \epsilon_2 ..., \epsilon_D]^T$.

Also take note that $\mathbb{E}[\epsilon_i] = 0$ and $\mathbb{E}[\epsilon_i \epsilon_j] = \delta_{ij}\sigma^2$.

Show that minimizing the noise-less error with L_2 regularization(excluding bias) yields the same result as optimizing the noise injected error (averaged over the noise distribution)

Let E'_d represent the expected error over the noisy input. Then we need to show that

$$\mathbb{E}_{\epsilon \sim \mathcal{N}(\mathbb{0}, \Sigma)}[E'_d] \sim E_d + \lambda \sum_{i=1}^{D} |w_i|^2$$

The initial design matrix can be given by a $d \times N$ *matrix* X and the actuals can be represented by an $N - vector$ Y

let X' be the design matrix after noising the inputs by ϵ. Then,

$$X' \triangleq X + \epsilon \mathbb{1}_{1\times N}$$

This is simply adding N-concantenated $\epsilon's$ to the design matrix (ϵ is a $d-vector$ i.e. $d \times 1$ in this expression).
Let $W \triangleq [w_1, w_2 ..., w_D]^T$

now the prediction vectors $\hat{Y}$ and $\hat{Y}'$ are as follows:

$$\begin{aligned}
\hat{Y} &= X^T W + w_0 \mathbb{1}_D \\
\hat{Y}' &= X'^T W + w_0 \mathbb{1}_D \\
&= (X + \epsilon \mathbb{1}_{1\times N})^T W + w_0 \mathbb{1}_D \\
&= (X^T + \mathbb{1}_{N\times 1}\epsilon^T) W + w_0 \mathbb{1}_D \\
&= \hat{Y} + \mathbb{1}_{N\times 1}\epsilon^T W
\end{aligned}$$

The loss expressions E_d and E'_d are as follows:

$$\begin{aligned}
E_d &= \frac{1}{2}||\hat{Y} - Y||_2^2 \\
E'_d &= \frac{1}{2}||\hat{Y}' - Y||_2^2
\end{aligned}$$

Now

$$\begin{aligned}
E'_d &= \frac{1}{2}(\hat{Y} + \mathbb{1}_{N\times 1}\epsilon^T W - Y)^T(\hat{Y} + \mathbb{1}_{N\times 1}\epsilon^T W - Y) \\
&= \frac{1}{2}(\hat{Y} - Y + \mathbb{1}_{N\times 1}\epsilon^T W)^T(\hat{Y} - Y + \mathbb{1}_{N\times 1}\epsilon^T W) \\
&= \frac{1}{2}(\hat{Y} - Y)^T + (\epsilon^T W)\mathbb{1}_{1\times N})((\hat{Y} - Y) + (\epsilon^T W)\mathbb{1}_{N\times 1}) \\
&= \frac{1}{2}\left((\hat{Y} - Y)^T(\hat{Y} - Y) + \epsilon^T W\left((\hat{Y} - Y)^T\mathbb{1}_{N\times 1} + \mathbb{1}_{1\times N}(\hat{Y} - Y)\right) + (\epsilon^T W)^2\mathbb{1}_{1\times N}\mathbb{1}_{N\times 1}\right) \\
&= E_d + \frac{1}{2}(\epsilon^T W)(2 \cdot (\hat{Y} - Y)^T\mathbb{1}_{N\times 1}) + \frac{1}{2}||\epsilon^T W||_2^2 \\
&= E_d + (\epsilon^T W)\alpha + \frac{1}{2}(\epsilon^T W)^T(\epsilon^T W) \\
&= E_d + (\epsilon^T W)\alpha + \frac{1}{2}(W^T \epsilon \epsilon^T W)
\end{aligned}$$

Now, the following holds true in regards to ϵ:

$$\mathbb{E}[\epsilon] = [\mathbb{E}[\epsilon_i]]^T = \mathbb{0}_{d\times 1}$$

$$let\ A_{d\times d} \triangleq \epsilon\epsilon^T$$

$$then\ A_{ij} = \epsilon_i\epsilon_j$$

Also note that each ϵ_i is independent to the others hence:

$$\begin{aligned} i \neq j \implies Cov(\epsilon_i, \epsilon_j) = \mathbb{E}[\epsilon_i\epsilon_j] - \mathbb{E}[\epsilon_i]\mathbb{E}[\epsilon_j] && = 0 \\ \therefore \forall i \neq j\ \mathbb{E}[\epsilon_i\epsilon_j] = \mathbb{E}[\epsilon_i]\mathbb{E}[\epsilon j] && = 0 \\ also/Var(\epsilon_i) = \mathbb{E}[\epsilon_i^2] - 0 && = \sigma^2 \end{aligned}$$

from this, it follows that

$$\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & otherwise \end{cases} \tag{1}$$

now finding expected A over ϵ:

$$\begin{aligned} \mathbb{E}_{\epsilon\sim\mathcal{N}(\mathbb{0},\Sigma)}[A]_{ij} &= \begin{cases} \sigma^2 & i = j \\ 0 & otherwise \end{cases} \\ \therefore \mathbb{E}_{\epsilon\sim\mathcal{N}(\mathbb{0},\Sigma)}[A]_{ij} &= \sigma^2\mathbb{I}_{d\times d} \end{aligned}$$

i.e. expectation of A over a ϵ is a diagonal matrix

From this, finally...

$$\begin{aligned} \mathbb{E}_{\epsilon\sim\mathcal{N}(\mathbb{0},\Sigma)}[E'_d] &= \mathbb{E}\left[E_d + (\epsilon^T W)\alpha + \frac{1}{2}(W^T A W)\right] \\ &= E_d + \mathbb{E}[\epsilon^T]W\alpha + \frac{1}{2}(W^T\mathbb{E}[A]W) \\ &= E_d + \mathbb{0}_{1\times d}W\alpha + \frac{1}{2}(W^T(\sigma^2\mathbb{I})W) \\ &= E_d + \frac{\sigma^2}{2}(W^T W) \\ &= E_d + \frac{\sigma^2}{2}\sum_{i=1}^{D}|w_i|^2 \end{aligned}$$

here λ is simulated by $\dfrac{\sigma^2}{2}$ ■