MAD Assignment 1

dmx289, Jakob Legaard

November 30, 2025

Contents

1 Exercise 1 (Weighted Average Loss)

1.1 a)

The weighted average loss is given by:

$$L = \frac{1}{N}\sum_{n=1}^{N} \alpha_n (w^T x_n - t_n)^2$$

- $\mathbf{X} \in \mathbb{R}^{N \cdot D}$ design matrix where each row is x_n^T
- $\mathbf{t} \in \mathbb{R}^N$ atarget vector $[t_1, t_2, \ldots, t_N]^T$
- $\mathbf{w} \in \mathbb{R}^D$ a parameter vector
- $\mathbf{A} \in \mathbb{R}^{N \cdot N}$ diagonal weight matrix with $A_{nn} = \alpha_n$

The loss function can be rewritten in matrix-vector form as:

$$L = \frac{1}{N}(\mathbf{Xw} - \mathbf{t})^T \mathbf{A} (\mathbf{Xw} - \mathbf{t})$$

expanding the matrix expression:

$$L = \frac{1}{N}[(\mathbf{Xw})^T \mathbf{A}(\mathbf{Xw}) - (\mathbf{Xw})^T \mathbf{At} - \mathbf{t}^T \mathbf{A}(\mathbf{Xw}) + \mathbf{t}^T \mathbf{At}]$$

$\mathbf{A}$ is symmetric and $(\mathbf{Xw})^T \mathbf{At}$ is a scalar equal to $\mathbf{t}^T \mathbf{A}(\mathbf{Xw})$:

$$L = \frac{1}{N}[\mathbf{w}^T \mathbf{X}^T \mathbf{AXw} - 2\mathbf{w}^T \mathbf{X}^T \mathbf{At} + \mathbf{t}^T \mathbf{At}]$$

gradient with respect to $\mathbf{w}$:

$$\nabla_{\mathbf{w}} L = \frac{1}{N}[2\mathbf{X}^T \mathbf{AXw} - 2\mathbf{X}^T \mathbf{At}]$$

using the following matrix identities:

$$\frac{\partial}{\partial \mathbf{w}}(\mathbf{w}^T \mathbf{Bw}) = 2\mathbf{Bw} for symmetric \mathbf{B}$$

$$\frac{\partial}{\partial \mathbf{w}}(\mathbf{w}^T \mathbf{c}) = \mathbf{c} \quad \textbf{for vector} \quad \mathbf{c}$$

we now set the gradient to zero and solve it:

$$\frac{1}{N}[2\mathbf{X}^T\mathbf{A}\mathbf{X}\mathbf{w} - 2\mathbf{X}^T\mathbf{A}\mathbf{t}] = 0$$

Simplifying it:

$$\mathbf{X}^T\mathbf{A}\mathbf{X}\mathbf{w} = \mathbf{X}^T\mathbf{A}\mathbf{t}$$

isolate w, by multiplying both sides by $(\mathbf{X}^T\mathbf{A}\mathbf{X})^{-1}$:

$$\hat{\mathbf{w}} = (\mathbf{X}^T\mathbf{A}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{A}\mathbf{t}$$

This is the optimal solution for weighted least squares regression, where $\mathbf{A} = \text{diag}(\alpha_1, \alpha_2, \dots, \alpha_N)$ is the diagonal matrix of weights.

1.2 b)

2 Exericise 2 Polynomial Fitting with Regularized Linear Regression and Cross-Validation

3 Exercise 3 Pdf and cdf

3.1 a)

we need to find the probability density function of

$$F(x) = \begin{cases} 0 & x \leq 0 \\ 1 - \exp(-\beta x^{\alpha}) & x > 0 \end{cases}$$

The PDF of this function is the derivative, so

$$pdf = f(x) = \frac{dF(x)}{dx}$$

FOr values of x>0 we have

$$F(x) = 1 - e^{-\beta x^{\alpha}}$$

Since $F(x) = 0$ for $x \leq 0$, we have:

$$f(x) = \frac{d}{dx}[0] = 0$$

We need to differentiate $F(x) = 1 - e^{(-\beta x^\alpha)}$:

$$\begin{aligned} f(x) &= \frac{d}{dx}\left[1 - e^{(-\beta x^\alpha)}\right] \\ &= 0 - \frac{d}{dx}\left[e^{(-\beta x^\alpha)}\right] \\ &= -e^{(-\beta x^\alpha)} \cdot \frac{d}{dx}[-\beta x^\alpha] \quad \text{(chain rule)} \\ &= -e^{(-\beta x^\alpha)} \cdot (-\beta \alpha x^{\alpha-1}) \quad \text{(power rule)} \\ &= \alpha\beta x^{\alpha-1} e^{(-\beta x^\alpha)} \end{aligned}$$

which gives us an pdf of the function F(x):

$$f(x) = \begin{cases} 0 & x \leq 0 \\ \alpha\beta x^{\alpha-1} e^{(-\beta x^\alpha)} & x > 0 \end{cases}$$

3.2 b)

We fix the parameters to $\alpha = 2$ and $\beta = \frac{1}{4}$. our cdf becomes:

$$F(x) = 1 - e^{\left(-\frac{x^2}{4}\right)} \quad \text{for } x > 0$$

We need to find $P(x > 4)$.

$$P(x > 4) = 1 - F(4)$$

insert:

$$\begin{aligned} P(x > 4) &= 1 - (1 - e^{-4}) = e^{-4} \\ &= 0.01831563888 \end{aligned}$$

So the probability is 1.83% approximately.

We now need to find the probability that the chip stops working in the time terminal $[5; 10]$. So

$$P(5 \leq x \leq 10)$$

which means

$$P(5 \leq x \leq 10) = P(10) - P(5)$$

The parameters are still fixed so we insert the values of x

$$P(10) - P(5) = (1 - e^{-25}) - (1 - e^{-25/4})$$

$$= -e^{-25} + e^{-25/4}$$

$$= 0.00193045415$$

SO the probability of a chips lifespan being in the interval of 5 to 10 years is approximately 0.193%

3.3 c)

We now need to find the median. The median m is the value where $F(m) = 0.5$.
Setting $F(m) = 0.5$

$$1 - e^{(-\beta m^{\alpha})} = 0.5$$

This means we solve for m:

$$1 - e^{(-\beta m^{\alpha})} = 0.5$$

$$e^{(-\beta m^{\alpha})} = 0.5$$

$$-\beta m^{\alpha} = \ln(0.5)$$

$$-\beta m^{\alpha} = -\ln(2)$$

$$\beta m^{\alpha} = \ln(2)$$

$$m^{\alpha} = \frac{\ln(2)}{\beta}$$

$$m = \left(\frac{\ln(2)}{\beta}\right)^{1/\alpha}$$

This means the median life span is $median = \left(\frac{\ln(2)}{\beta}\right)^{1/\alpha}$

4 Exercise4 Conditional probability and expectations

4.1 a)

We need to calculate the outcomes of remaining silent or not when arrested by the police. The crime for which the police have arrested the suspect is punishable by 5 years in prison. in days, this is equal to

$$5 \cdot 365 = 1825 \quad \textbf{days}$$

We have two types of people: those who have never been convicted and those who have been convicted before. ANd they have two choices: remaining silent or talking to the police during the arrest. We are given some probabilities:

$$P(\textbf{Court}|\textbf{Silent}) = 0.001$$

$$P(\textbf{Court}|\textbf{NC talks}) = 0.0015$$

$$P(\textbf{Court}|\textbf{C talks}) = 0.005$$

then we have probabilities of the outcome if the case goes to court:

$$P(\textbf{Acquittal}|\textbf{NC talked}) = 0.8$$

$$P(\textbf{Acquittal}|\textbf{C talked}) = 0.2$$

$$P(\textbf{Acquittal}|\textbf{NC no talk}) = 0.2$$

$$P(\textbf{Acquittal}|\textbf{C no talk}) = 0.05$$

And the last thing is, if the person talked at the arrest stage and gets convicted, the sentence is reduced by 50%.

4.1.1 Remain silent

The probabilities for remaining silent:
Case does not go to court: $P(\textbf{No Court}) = 1 - 0.001 = 0.999 \Rightarrow 0$ days
Case goes to court: $P(\textbf{Court}) = 0.001$

Acquitted: $P(\textbf{Acquittal} \mid \textbf{Court, Silent}) = 0.2 \Rightarrow 0$ days
Convicted: $P(\textbf{Conviction} \mid \textbf{Court, Silent}) = 1 - 0.2 = 0.8 \Rightarrow 1825$ days

The expected number of days in prison is:

$$\begin{aligned}
E[\textbf{Days} \mid \textbf{Silent, NC}] &= P(\textbf{No Court}) \cdot 0 \\
&\quad + P(\textbf{Court}) \cdot P(\textbf{Acquittal} \mid \textbf{Court, Silent}) \cdot 0 \\
&\quad + P(\textbf{Court}) \cdot P(\textbf{Conviction} \mid \textbf{Court, Silent}) \cdot 1825 \\
&= 0.999 \cdot 0 + 0.001 \cdot 0.2 \cdot 0 + 0.001 \cdot 0.8 \cdot 1825 \\
&= 0 + 0 + 1.46 \\
&= 1.46 \text{ days}
\end{aligned}$$

4.1.2 talks to police

The probabilities for talking to the police:
Case does not go to court: $P(\textbf{No Court}) = 1 - 0.0015 = 0.9985 \Rightarrow 0$ days
Case goes to court: $P(\textbf{Court}) = 0.0015$

Acquitted: $P(\textbf{Acquittal} \mid \textbf{Court, Talk}) = 0.8 \Rightarrow 0$ days
Convicted: $P(\textbf{Conviction} \mid \textbf{Court, Talk}) = 1 - 0.8 = 0.2 \Rightarrow 1825 \cdot 0.5 = 912.5$ days

$$\begin{aligned}
E[\textbf{Days} \mid \textbf{Talk, NC}] &= P(\textbf{No Court}) \cdot 0 \\
&\quad + P(\textbf{Court}) \cdot P(\textbf{Acquittal} \mid \textbf{Court, Talk}) \cdot 0 \\
&\quad + P(\textbf{Court}) \cdot P(\textbf{Conviction} \mid \textbf{Court, Talk}) \cdot 912.5 \\
&= 0.9985 \cdot 0 + 0.0015 \cdot 0.8 \cdot 0 + 0.0015 \cdot 0.2 \cdot 912.5 \\
&= 0 + 0 + 0.27375 \\
&= 0.27 \text{ days}
\end{aligned}$$

This means that Talking to the police during the arrest would result in an expected lower amount of days in prison.

4.2 b)

Now we do the same, but with the history of convictions.

4.2.1 Remain silent

The probabilities for remaining silent: Case does not go to court $P(\textbf{No Court}) = 1 - 0.001 = 0.999$ $\Rightarrow 0$ days
Case goes to court: $P(\textbf{Court}) = 0.001$

Acquitted: $P(\textbf{Acquittal}|\textbf{court,Silent}) = 0.05 \Rightarrow 0$ days
Convicted: $P(\textbf{Convicted}|\textbf{Court,silent}) = 1 - 0.05 = 0.95 \Rightarrow 1825$days

$$\begin{aligned}
E[\textbf{Days} \mid \textbf{Silent, C}] &= P(\textbf{No Court}) \cdot 0 \\
&\quad + P(\textbf{Court}) \cdot P(\textbf{Acquittal} \mid \textbf{Court, Silent}) \cdot 0 \\
&\quad + P(\textbf{Court}) \cdot P(\textbf{Conviction} \mid \textbf{Court, Silent}) \cdot 1825 \\
&= 0.999 \cdot 0 + 0.001 \cdot 0.05 \cdot 0 + 0.001 \cdot 0.95 \cdot 1825 \\
&= 0 + 0 + 1.73375 \\
&= 1.73 \text{ days}
\end{aligned}$$

4.2.2 Talks to police

The probabilities for remaining silent:
Case does not go to court $P(\textbf{No Court}) = 1 - 0.005 = 0.995 \Rightarrow 0$ days
Case goes to court: $P(\textbf{Court}) = 0.005$

Acquitted: $P(\textbf{Acquittal}|\textbf{court,talks}) = 0.2 \Rightarrow 0$ days
Convicted: $P(\textbf{Convicted}|\textbf{Court,talks}) = 1 - 0.2 = 0.8 \Rightarrow 1825 \cdot 0.5 = 912.5$days

$$\begin{aligned}
E[\textbf{Days} \mid \textbf{Talk, C}] &= P(\textbf{No Court}) \cdot 0 \\
&\quad + P(\textbf{Court}) \cdot P(\textbf{Acquittal} \mid \textbf{Court, Talk}) \cdot 0 \\
&\quad + P(\textbf{Court}) \cdot P(\textbf{Conviction} \mid \textbf{Court, Talk}) \cdot 912.5 \\
&= 0.995 \cdot 0 + 0.005 \cdot 0.2 \cdot 0 + 0.005 \cdot 0.8 \cdot 912.5 \\
&= 0 + 0 + 3.65 \\
&= 3.65 \text{ days}
\end{aligned}$$

this means that a person with a conviction history has better odds by remaining silence when arrested by the police, as $E[\textbf{Days} \mid \textbf{Talk, C}] > E[\textbf{Days} \mid \textbf{Silent, C}]$