MAD 2022-23, Assignment 2

Helga Rykov Ibsen <mcv462> Hold 6

1. september 2023

Opgave 1

(a)

Vi skal finde de optimale estimater $\hat{w}_0$ og $\hat{w}_1$ for den *vægtede gennemsnitligt tab* $\mathcal{L}$:

$$\mathcal{L} = \frac{1}{N}\sum_{n=1}^{N}\alpha_n(f(\mathbf{x}_n;\mathbf{w}) - t_n)^2 = \frac{1}{N}\sum_{n=1}^{N}\alpha_n(\mathbf{w}^T\mathbf{x}_n - t_n)^2$$

Vi starter med at omskrive det første led i parentesen i det sidste udtryk på formen af ligningen på koefficientformen:

$$\begin{aligned}\mathcal{L} &= \frac{1}{N}\sum_{n=1}^{N}\alpha_n(\mathbf{w}^T\mathbf{x}_n - t_n)^2 \\ &= \frac{1}{N}\sum_{n=1}^{N}\alpha_n(w_0 + w_1x_n - t_n)^2 \\ &= \frac{1}{N}\sum_{n=1}^{N}\alpha_n(w_0^2 + w_1^2x_n^2 + t_n^2 + 2w_0w_1x_n - 2w_1t_nx_n - 2w_0t_n)\end{aligned}$$

Vi skal nu differentiere udtrykket mht. variablerne $\hat{w}_0$ og $\hat{w}_1$:

$$\frac{\partial\mathcal{L}}{\partial w_0} = \frac{1}{N}\sum_{n=1}^{N}\alpha_n(2w_0 + 2w_1x_n - 2t_n)$$

$$\frac{\partial\mathcal{L}}{\partial w_1} = \frac{1}{N}\sum_{n=1}^{N}\alpha_n(2w_1x_n^2 + 2w_0x_n - 2t_nx_n)$$

Vi ganger α_n og $\frac{1}{N}$ ind i parentesen:

$$\frac{\partial\mathcal{L}}{\partial w_0} = \frac{1}{N}\sum_{n=1}^{N}\alpha_n2w_0 + \frac{1}{N}\sum_{n=1}^{N}2w_1\alpha_nx_n - \frac{1}{N}\sum_{n=1}^{N}2\alpha_nt_n$$

$$\frac{\partial\mathcal{L}}{\partial w_1} = \frac{1}{N}\sum_{n=1}^{N}\alpha_n2w_1\alpha_nx_n^2 + \frac{1}{N}\sum_{n=1}^{N}2w_0\alpha_nx_n - \frac{1}{N}\sum_{n=1}^{N}2\alpha_nt_nx_n$$

Nu skal vi løse ligningssystemet for $\hat{w}_0$ og $\hat{w}_1$. Vi sætter de to ligninger lige med 0, smider det sidste led på den anden side af lighedstegnet og dividerer med 2 på begge sider. $\frac{1}{N}$ og summationstegnene går ud ad de udligner hinanden. n er gemt i gennemsnittet og kan også bortforkortes:

$$\frac{\partial \mathcal{L}}{\partial w_0} = 0 \Leftrightarrow w_0\overline{\alpha} + w_1 \cdot \overline{\alpha x} = \overline{\alpha t}$$

$$\frac{\partial \mathcal{L}}{\partial w_1} = 0 \Leftrightarrow w_1\overline{\alpha x^2} + w_0 \cdot \overline{\alpha x} = \overline{\alpha t x}$$

Vi starter med at isolere w_0 i den første ligning og bestemme udtrykket for det første optimale estimatet $\hat{w}_0$:

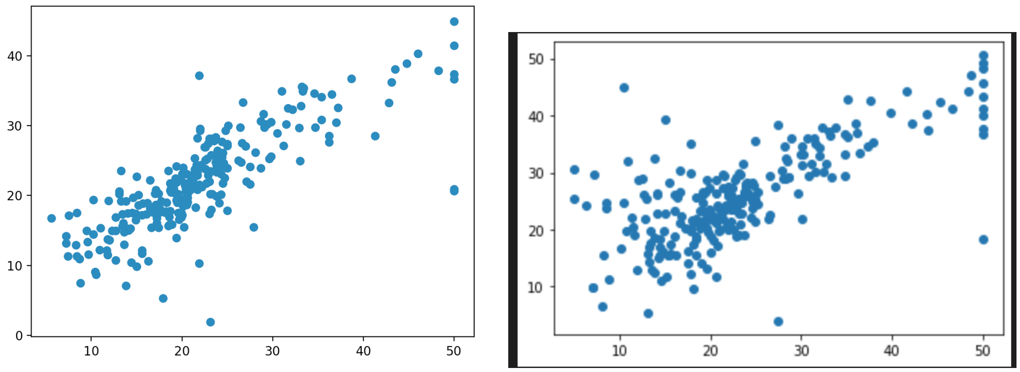
$$\begin{aligned}
\hat{w}_0\overline{\alpha} &= \overline{\alpha t} - w_1\overline{\alpha x} \\
\hat{w}_0 &= \frac{1}{\overline{\alpha}} \cdot (\overline{\alpha t} - w_1\overline{\alpha x})
\end{aligned}$$

For at finde estimatet $\hat{w}_1$ indsætter vi det fundne $\hat{w}_0$:

$$\begin{aligned}
\hat{w}_1\overline{\alpha x^2} + \frac{1}{\overline{\alpha}}(\overline{\alpha t} - \hat{w}_1\overline{\alpha x})\overline{\alpha x} &= \overline{\alpha t x} \\
\hat{w}_1\overline{\alpha x^2} + \frac{\overline{\alpha t}}{\overline{\alpha}}\overline{\alpha x} - \hat{w}_1\frac{(\overline{\alpha x})^2}{\overline{\alpha}} &= \overline{\alpha t x} \\
\hat{w}_1(\overline{\alpha x^2} - \frac{(\overline{\alpha x})^2}{\overline{\alpha}}) &= \overline{\alpha t x} - \frac{\overline{\alpha t} \cdot \overline{\alpha x}}{\overline{\alpha}} \\
\hat{w}_1 &= \frac{\overline{\alpha} \cdot \overline{\alpha t x} - \overline{\alpha t} \cdot \overline{\alpha x}}{\overline{\alpha} \cdot \overline{\alpha x^2} - \overline{\alpha x}^2}
\end{aligned}$$

(b)

Ved at bruge det vægtede gennemsnit forventer jeg at få en model som kan forudse udviklingen af mine data bedre end den simple lineære model.
Det er imidlertid ikke hvad jeg observerer ved at kigge på scatterplot'tet for det vægtede gennemsnit (b) i Figur 1 nedenunder. Datapunkterne i (b) ligger mere spredt end i (a) og modellen (b) har dermed større spredning end (a). Umiddelbart, gør dette modellen (b) svagere end (a), hvilket burde jo være omvendt.



(a) (b)

Figur 1: The 2D scatter plots of (a) the linear model based on all features and (b) the non-linear model based on all features.

Opgave 3

(a)

Vi er givet en fordelingsfunktion (cdf):

$$F(x) = 1 - e^{-\beta x^{\alpha}} \quad \text{x>0} \tag{1}$$

Vi skal bestemme tæthedsfunktion (pdf). Vi ved at fordelingsfunktion er en stamfunktion til tæthedsfunktion og finder derfor pdf'en ved at differentiere cdf'en:

$$\begin{aligned} F' = f(x) &= (1 - e^{-\beta x^{\alpha}})' \\ &= -e^{-\beta x^{\alpha}} \cdot (-\beta x^{\alpha})' \\ &= \beta \alpha x^{\alpha - 1} \cdot e^{-\beta x^{\alpha}} \end{aligned}$$

(b)

For at finde sandsynlighed for at chipene virker længere end 4 år, skal vi bruge cdf'en P(x < 4) og trække den fra 1:

$$\begin{aligned}
P(x > 4) &= 1 - P(x < 4) \\
&= 1 - F(4) = 1 - (1 - e^{-\beta x^{\alpha}}) \qquad \alpha{=}2, \quad \beta{=}1/4 \\
&= e^{-1/4 \cdot 4^2} \\
&= e^{-4} = 0.0183 \qquad \text{(ca. \%2)}
\end{aligned}$$

For at finde sandsynlighed for at chipene virker i intervallet [5; 10] år, skal vi bruge cdf'en P(5 < x < 10):

$$\begin{aligned}
P(5 < x < 10) &= P(10) - P(5) \\
&= (1 - e^{-1/4 \cdot 10^2}) - (1 - e^{-1/4 \cdot 5^2}) \qquad \alpha{=}2, \quad \beta{=}1/4 \\
&= -e^{-1/4 \cdot 10^2} + e^{-1/4 \cdot 5^2} \\
&= -e^{-25} + e^{-6.25} = 0.0019 \qquad \text{(ca. 0.2 \%)}
\end{aligned}$$

(c)

For fordelinger hvis pdf er symmetriske svarer middelværdien til symmetriaksen og dermed medianen m. Vi skal derfor finde medianen:

$$\begin{aligned}
F(m) &= 0.5 \\
1 - e^{-\beta m^{\alpha}} &= 0.5 \\
0.5 &= e^{-\beta m^{\alpha}} \\
ln(0.5) &= -\beta m^{\alpha} \\
\left(\frac{ln(0.5)}{-\beta}\right)^{1/\alpha} &= m
\end{aligned}$$

Opgave 4

(a)

Vi anvender formlen for gennemsnitsværdien for variablen X (X = 5 års betinget fængsel):

$$E_{P(x)}\{X\} = \sum_x xP(x) \tag{2}$$

1. Case: person NC taler til politiet:

$$\begin{aligned}E_{P(5)}\{5\} &= \sum 5P(5)\\ &= 5 \cdot (P(\text{court} \mid \text{NC}) = 0.002) +\\ &+ 5 \cdot (P(\text{NC prisoned} \mid \text{court}) = 0.5) =\\ &= 0.01 + 2.5 = 2.51\end{aligned}$$

   Vi ved at hvis personen er blevet dømt OG har talt med politiet, så bliver dens dom reduceret med faktor 0.75, derfor kommer hans forventede fængselsdom til at være:

$$E_{P(5)}\{5\} \cdot 0.75 = 2.51 \cdot 0.75 = 1.882 \quad \text{(ca. 2 år)}$$

2. Case: NC taler ikke til politiet:

$$\begin{aligned}E_{P(5)}\{5\} &= \sum 5P(5)\\ &= 5 \cdot (P(\text{court}) = 0.001) + 5 \cdot (P(NC\text{prisoned} \mid \text{court}) = 0.5)\\ &+ 5 \cdot (P(NC(\text{not prisoned}|\text{not talk to police}) = 0.5/4) =\\ &= 0.005 + 2.5 + 0.625 = 3.13 \quad \text{(ca. 3 år)}\end{aligned}$$

(b)

Vi anvender formlen (1) for den gennemsnitlige værdi.

1. Case: person C taler til politiet:

$$\begin{aligned}E_{P(5)}\{5\} &= \sum 5P(5)\\ &= 5 \cdot (P(\text{court} \mid \text{C talked to police}) = 0.005) +\\ &+ 5 \cdot (P(\text{C prisoned} \mid \text{court}) = 1 - 0.1) =\\ &= 0.025 + 4.5 = 4.525\end{aligned}$$

   Vi ved at hvis personen er blevet dømt OG har talt med politiet, så bliver dens dom reduceret med faktor 0.75, derfor kommer hans forventede fængselsdom til at være:

$$E_{P(5)}\{5\} \cdot 0.75 = 4.525 \cdot 0.75 = 3.393 \quad \text{(ca. 3.4 år)}$$

2. Case: person C taler ikke til politiet:

$$\begin{aligned}
E_{P(5)}\{5\} &= \sum 5P(5) \\
&= 5 \cdot (P(\text{court}) = 0.001) + 5 \cdot (P(C\text{prisoned} \mid \text{court}) = 1 - 0.1) \\
&+ 5 \cdot (P(C(\text{not prisoned}|\text{not talk to police}) = 0.1/4) = \\
&= 0.005 + 4.5 + 0.125 = 4.63 \quad \text{(ca. 4.6 år)}
\end{aligned}$$