Notes Date / Dec / 2025

WiDS

W-1

3.1

1\.
2\. $\begin{bmatrix} 4 & 1 \\ 1 & 3 \end{bmatrix} w = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ $A^{-1} = \frac{1}{|A|}\begin{bmatrix} +3 & -1 \\ -1 & +4 \end{bmatrix}^T = \frac{1}{|A|}\begin{bmatrix} 3 & -1 \\ -1 & 4 \end{bmatrix}$

$w = A^{-1} b$

$w = \frac{1}{11}\begin{bmatrix} 3 & -1 \\ -1 & 4 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \frac{1}{11}\begin{bmatrix} 1 \\ 7 \end{bmatrix}$

3\. $\frac{1}{11}\begin{bmatrix} 4 & 1 \\ 1 & 3 \end{bmatrix}\begin{bmatrix} 1 \\ 7 \end{bmatrix} = \frac{1}{11}\begin{bmatrix} 11 \\ 22 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$

3.2

1\. $2x+3$

2\. $\frac{\delta g}{\delta x_1} = 2x_1 + 2x_2$ $\frac{\delta g}{\delta x_2} = 2x_1 + 6x_2$

3\. $\begin{bmatrix} 2x \\ 3 \end{bmatrix}$

4\. $\begin{bmatrix} \frac{\delta g}{\delta x_1} \\ \frac{\delta g}{\delta x_2} \end{bmatrix} = \nabla_x g = \begin{bmatrix} 2x_1 + 2x_2 \\ 2x_1 + 6x_2 \end{bmatrix}$

Notes

Date / /

5. $g(x_1, x_2) = \cancel{x x^T A}\ x A x^T$

$= \cancel{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} [x_1 \ x_2] A}$

$\cancel{x_1^2 + 2x_1x_2 + 3x_2^2 = (x_1^2 + x_2^2) A}$

$\cancel{A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}}$

$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$

$g(x_1, x_2) = x_1^2 + 2x_1x_2 + 3x_2^2 = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} [x_1 \ x_2]$

$A = \begin{bmatrix} 1 & 1 \\ 1 & 3 \end{bmatrix}$

$= \begin{bmatrix} ax_1 + bx_2 \\ cx_1 + dx_2 \end{bmatrix} [x_1 \ x_2]$

$= ax_1^2 + bx_1x_2 + cx_1x_2 + dx_2^2$

$x_1^2 + 2x_1x_2 + 3x_2^2 \qquad = ax_1^2 + (b+c)x_1x_2 + dx_2^2$

$a = 1 \qquad d = 3 \qquad b + c = 2$

$b = c = 1$

6. $\frac{\delta}{\delta x}(x^T A x) = \frac{\delta}{\delta x}(g(x)) = \nabla_x g = \begin{bmatrix} 2x_1 + 2x_2 \\ 2x_1 + 6x_2 \end{bmatrix}$

$\nabla_x g = px$

$= p \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2x_1 + 2x_2 \\ 2x_1 + 6x_2 \end{bmatrix}$

$p = 2A$

$2 \times \begin{bmatrix} 1 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} x_1 & x_2 \\ x_2 & \end{bmatrix}$

Notes

Date / /

7. (a) $A^T x = \cancel{Ax}$

$= \sum a_i x_i$

$\frac{\delta}{\delta x}(A^T x) = \begin{bmatrix} \frac{\delta A^T x}{\delta x_1} \\ \frac{\delta A^T x}{\delta x_2} \end{bmatrix}$

$A^T = \begin{bmatrix} a_1 & a_3 \\ a_2 & a_4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

$A^T x = a_1 x_1 + a_3 x_2 + a_2 x_1 + a_4 x_2$

$\frac{\delta f}{\delta x} = \begin{bmatrix} \delta f / \delta x_1 \\ \delta f / \delta x_2 \end{bmatrix}$ $\frac{\delta}{\delta x}[A^T x] = \begin{bmatrix} a_1 + a_2 \\ a_3 + a_4 \end{bmatrix}$

$= A$

$\frac{\delta f}{\delta x_k}\left(\sum a_k x_k \right) = \sum a_k = A$

(b) $\cancel{x} = \begin{bmatrix} \cancel{x_1} \\ x_2 \end{bmatrix}$

$A = \begin{bmatrix} a_1 & \cancel{a_2} \\ a_2 & \cancel{a_4} \\ a_3 \\ \vdots \\ a_n \end{bmatrix}$ $x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix}$

$x^T x = [x_1 \ x_2 \ldots x_n] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix}$

$= x_1^2 + x_2^2 + \ldots x_n^2$

$\frac{\delta}{\delta x}(x^T x) = \begin{bmatrix} \frac{\delta}{\delta x_1}(x^T x) \\ \frac{\delta}{\delta x_2}(x^T x) \\ \vdots \end{bmatrix} = \begin{bmatrix} 2x_1 \\ 2x_2 \\ \vdots \\ 2x_n \end{bmatrix} = 2x$

(c) $[x_1 \ x_2 \ldots x_n] \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = a_1 x_1 + \ldots a_n x_n$

$[a_1 x_1 + \ldots a_n x_n] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$

$= \cancel{a_1 x_1^2 + a_2 x_2^2 + \ldots} a_n x_n^2$

$\frac{\delta}{\delta x} \cancel{(x^T A x)} = \begin{matrix} 2a_1 x_1 \\ 2a_2 x_2 \\ \vdots \end{matrix}$

$a_1 x_1^2 + \ldots a_n x_n$

$=$

$x^T A x \Rightarrow$ vector $a_1 x_1^2 \; a_2 x_2^2 \ldots a_n x_n$

Notes

Date / /

$A \neq \begin{bmatrix} a_1 \\ a_2 \\ \vdots \end{bmatrix}$ $A:$ $a_{11}\ a_{12} \cdots$

$a_{21}\ a_{22}$

$x^T A = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \cdots \\ a_{21} & a_{22} & \end{bmatrix}$

$x^T A x$ $\begin{bmatrix} a_{11}x_1 + a_{21}x_2 + \cdots a_{n1}x_n & a_{12}x_1 + a_{22}x_2 + \cdots a_{n2}x_n \end{bmatrix} \cdots \begin{bmatrix} x_1 \\ x_2 \\ x_n \end{bmatrix}$

$a_{11}x_1^2 + a_{21}x_1x_2 + \cdots a_{n1}x_1x_n$

$+ a_{12}x_1x_2 + a_{22}x_2^2 + \cdots a_{n2}x_2x_n$

$+ \cdots a_{1n}x_1x_n + a_{2n}x_2x_n + \cdots a_{nn}x_n^2$

$\frac{\delta}{\delta x} = a_{11}x_1 + a_{21}x_2 + \cdots a_{n1}x_n$

$+ a_{12}x_2 + a_1 \cdots a_{1n}$

$a_{1n}x_n$

$x^T A x = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} x_i x_j$

$\frac{\delta}{\delta x_k}(x^T A x) = \sum_{j=1}^{n} A_{kj} x_j + \sum_{i=1}^{n} A_{ik} x_i$

$i = k \Rightarrow x_j$

$j = k \Rightarrow x_i$

Ax $\begin{matrix} a_{k1} & \cdots & a_{kn} \end{matrix}$ $\begin{matrix} x_1 \\ \vdots \\ x_n \end{matrix}$

$(Ax)_k$

a_{1k}

$= (A + A^T)x$ $(A^T x)_k$ a_{2k}

Notes Date / /

4.1 $P(D|P) = \frac{\cancel{P(P|D)}P(D)P(P|D)\cancel{P(P\cap D)}P(D)}{P(P|D)+P(P|\bar{D})}$

$\frac{0.01\times0.99}{0.99+0.05} = \frac{0.01}{1.04} = \frac{1}{104}$

$\frac{P(P|D)\,P(D)}{P(P|D)+P(P|\bar{D})} = \frac{0.99\times0.01}{0.99+0.05}$

$= \frac{99\times10^{-4}}{104\times10^{-2}} = \frac{99}{10400}$

~~4.2~~ 5.

1. 10

2. Simple generality of $f(x)=2x$

3
$8a+4b+2c+d=4$ $\quad 16a+8b+4c+2d=$
$64a+16b+4c+d=8$ $\quad 64a+16b+4c+d$
$729a+81b+9c+d=18$ $\quad d=48a+8b$

3. No

4. Assumption that the relation is linear

5. Single input variable x has ~~parameters~~ features $x_1, x_2, \ldots x_n$ which are linearly related to y.

6. ~~$x_1 x_2$~~ $a_1x_1+a_2x_2+\ldots a_dx_d = y$
$+e$

Notes Date / /

W-2

1. $p(x) = \dfrac{e^{B_0 + B_1 x_1 + B_2 x_2}}{1 + e^{B_0 + B_1 x_1 + B_2 x_2}}$

$p(x) + p(x) e^{B_0 + B_1 x_1 + B_2 x_2} = e^{B_0 + B_1 x_1 + B_2 x_2}$

$e^{B_0 + B_1 x_1 + B_2 x_2} (1 - p(x)) = p(x)$

$\log\left(\dfrac{p(x)}{1 - p(x)}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$

$p_i = P(y_i = 1 \mid x_{1i}, x_{2i})$

(a) $\log\left(\dfrac{p_i}{1 - p_i}\right) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}$

$\Rightarrow$ Similar for each i

(b) $P(y_i = y_i) = p_i^{y_i} (1 - p_i)^{1 - y_i}$

$y_i = 1 / 0$

$P(y_i = 1) = p_i^{1} (1 - p_i)^{0}$ $\quad P(y_i = 0) = p_i^{0} (1 - p_i)^{1}$

Likelihood $= \prod_{i=1}^{n_1 + n_2} p_i = \prod_{i=1}^{n_1} P(y_i = 1) \prod_{i=n_1}^{n_1 + n_2} P(y_i = 0)$

$= \prod_{i=1}^{n_1 + n_2} p_i^{y_i} (1 - p_i)^{1 - y_i}$

$n_1 \Rightarrow p_i \ (1)$ $\quad n_2 \Rightarrow 1 - p_i \ (0)$

Notes Date / /

Log likelihood $= \prod_{i=1}^{n_1+n_2} \log\left[p_i^{y_i}(1-p_i)^{1-y_i}\right]$

$= \prod_{i=1}^{n_1+n_2} (y_i \log p_i + (1-y_i)\log(1-p_i))$

2. $\log\left(\frac{p_i}{1-p_i}\right) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$

$= -6 + 0.05x_1 + x_2$

(A) $\log\left(\frac{p_i}{1-p_i}\right) = -6 + 0.05(40) + 3.5$

$= 2 - 6 + 3.5 = -0.5$

$\log\left(\frac{1-p_i}{p_i}\right) = 0.5$

$\frac{1-p_i}{p_i} = e^{1/2}$ $\quad e^{\frac{1}{2}} p_i = 1 - p_i$

$p_i(1+e^{1/2}) = 1$

$p_i = \frac{1}{1+e^{\frac{1}{2}}}$

(B) $\log\left(\frac{\frac{1}{2}}{1-\frac{1}{2}}\right) = -6 + 0.05(x_1) + 3.5$

$\log 1 = -2.5 + 0.05x_1$

$\frac{2.5}{0.05} = x_1 = 50$

Date / /

Notes

3. $$P(y=1|x=4) = \frac{f_1(x)P(y=1)}{f_1(x)P(y=1)+f_0(x)P(y=0)}$$

$$f_k(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-1}{2\sigma^2}(x-\mu)^2\right)$$

$$P(y=1|x=4) = \frac{\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(\frac{-1}{2\sigma^2}(x-10)^2\right)\times\frac{8}{10}}{\frac{8}{10}\times\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(\frac{-1}{2\sigma^2}(x-10)^2\right)+\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(\frac{-1}{2\sigma^2}(x-0)^2\right)\times\frac{2}{10}}$$

$$= \frac{8\exp\left(\frac{-1}{2\times36}\times6^2\right)}{8\exp\left(\frac{-1}{2\times36}\times6^2\right)+2\exp\left(\frac{-1}{\cancel{2}\times\cancel{36}\,9}\times\cancel{4^2}\,\cancel{8}\,2\right)}$$

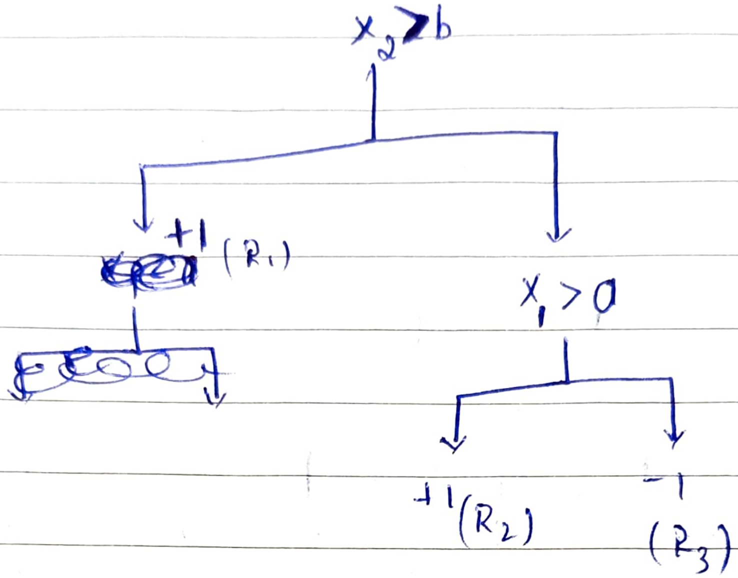
$$= \frac{8\exp\left(\frac{-1}{2}\right)}{8\exp\left(\frac{-1}{2}\right)+2\exp\left(\frac{-2}{9}\right)}$$

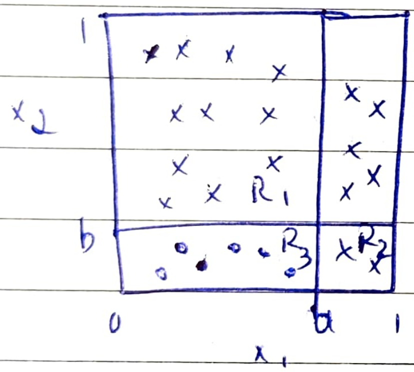
$$= \frac{4e^{-1/2}}{4e^{-1/2}+e^{\frac{-2}{9}}}$$

Notes Date / /

W-3

1





2. Random forest is a method that uses the average of predictions from many decision trees in order to predict the output by reducing the RSE and uses the maximum of the predictions in classification algorithms. It is an ensemble method. Forest essentially refers that many decision trees are used and the predictors are taken randomly p. The numbers of predictors

Notes Date / /

used is the square root of the total no. of predictors. Unlike bagging which takes only 1 predictor in consideration, this randomness ensures that the many predictions do not have much correlation which can happen in bagging. Hence overall, it reduces the error and improves performance.

3. Ensemble method in machine learning is basically considering many models which give different predictions and averaging the results (in regression) or choosing the most repeated classes (in classification) in order to overcome the narrowness of a single model. Combining multiple decision trees to produce the most relevant output is an ensemble method as it reduces the error by taking into consideration multiple possible decision trees rather than a single one. It can be done using bagging, random forests, etc.

Notes Date / /

4.1 TP : 180 FP : 70 TN : 730 FN : 20

2. Accuracy = $\frac{\text{Correct}}{\text{All}} = \frac{TP + TN}{TP+TN+FP+FN}$

$= \frac{180 + 730}{1000} = \frac{910}{1000} = 0.91$

Precision = $\frac{\text{~~Actua~~ Predicted correct positive}}{\text{All ~~Actual~~ predicted positives}}$

$= \frac{TP}{TP+FP} = \frac{180}{180+70} = \frac{180}{250} = 0.72$

Recall / Sensitivity = $\frac{\text{Predicted correct +}}{\text{Actual +}}$

$= \frac{TP}{TP+FN} = \frac{180}{180+20} = \frac{180}{200} = 0.9$

Specificity = $\frac{\text{Predicted correct -}}{\text{Actual -}} = \frac{TN}{TN+FP}$

$= \frac{730}{730+70} = \frac{730}{800} = 0.9125$

F1 Score = $\frac{2}{\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}} = \frac{2}{\frac{250}{180} + \frac{200}{180}} = \frac{2 \times 180}{450} = 0.8$

Notes Date / /

3. Specificity as gives importance to -vely predicted outputs which are correct.

4. If classification threshold is lowered, more no. of +ve predictions are possible, and negative predictions reduce.

TP & FP ↑

5. Yeah, accuracy is just the fraction of correct predictions, but it does not take into account ~~which~~ how many + and ~~wht-r~~ predictions are correct. Two classifiers with same accuracy can have different recall and precision.

Ex. TP = 180 FP = 70 TN = 730 FN = 20

Accuracy = 0.91 Precision = $\frac{180}{250}$ = 0.72

Recall = $\frac{180}{200}$ = 0.9

TP = 100 FP = 20 TN = 810 FN = 70

Accuracy = 0.91 Precision = $\frac{100}{120}$ = 0.83

Recall = $\frac{100}{170}$ = 0.58

Notes

Date / /

The WiDS project has really been entertaining and a really new learning opportunity for me till as this is my first time learning about and working with ML. Starting from the basics of math to linear regression, classification methods, logistic regression and decision trees, it had gone very good, but not very easy also. I frankly accept that I had to skip various contents of the book as I really couldn't understand anything. And in the beginning I haven't completed some parts of the assignment and have skipped the optional kaggle challenges due to other projects, but I hope to do them ~~if~~ whenever I have time. Overall it has been a great experience till now, thanks to my project mentors' guidance.