Notes

Date / /

W-2

1. $$p(x) = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2}}$$

$$p(x) + p(x)\, e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2} = e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2}$$

$$e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2} (1 - p(x)) = p(x)$$

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

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

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

$\Rightarrow$ similar for each i

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

$y_i = 1 / 0$

$$P(Y_i = 1) = p_i^1 (1 - p_i)^0 \qquad P(Y_i = 0) = p_i^0 (1 - p_i)^1$$

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

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

Notes Date / /

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

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

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

$$= -6 + 0.05 x_1 + x_2$$

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

$$= 2 - 6 + 3.5 = -0.5$$

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

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

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

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

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

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

$$\log 1 = -2.5 + 0.05 x_1$$

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

Notes

Date / /

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

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

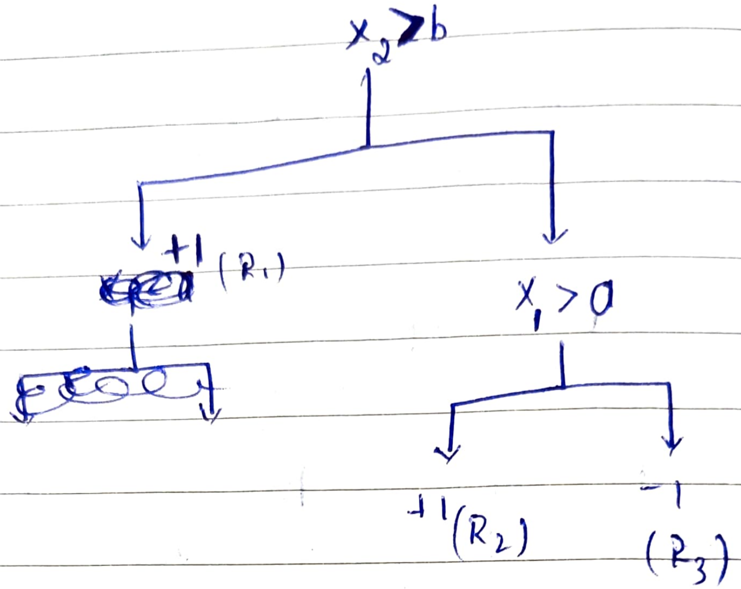
$$P(y=1 \mid x=4) = \frac{\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-1}{2\sigma^2}(x-10)^2\right) \times \frac{8}{10}}{\frac{8}{10} \times \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-1}{2\sigma^2}(x-10)^2\right) + \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-1}{2\sigma^2}(x-0)^2\right) \times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

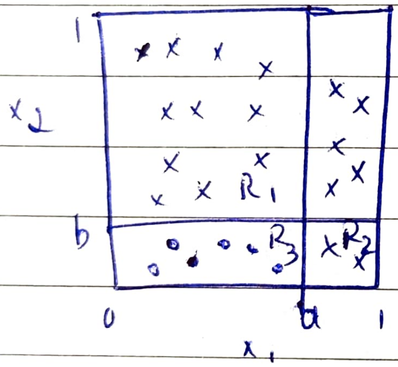
$$= \frac{8\exp\left(\frac{-1}{2\times 36} \times 6^2\right)}{8\exp\left(\frac{-1}{2\times 36} \times 6^2\right) + 2\exp\left(\frac{-1}{\cancel{2\times 36}\,9} \times \cancel{4^2}\,\cancel{8}\,2\right)}$$

$$= \frac{8\exp\left(\frac{-1}{2}\right)}{8\exp\left(\frac{-1}{2}\right) + 2\exp\left(\frac{-2}{9}\right)} = \frac{4e^{-1/2}}{4e^{-1/2} + e^{\frac{-2}{9}}}$$

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 ~~what~~ 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