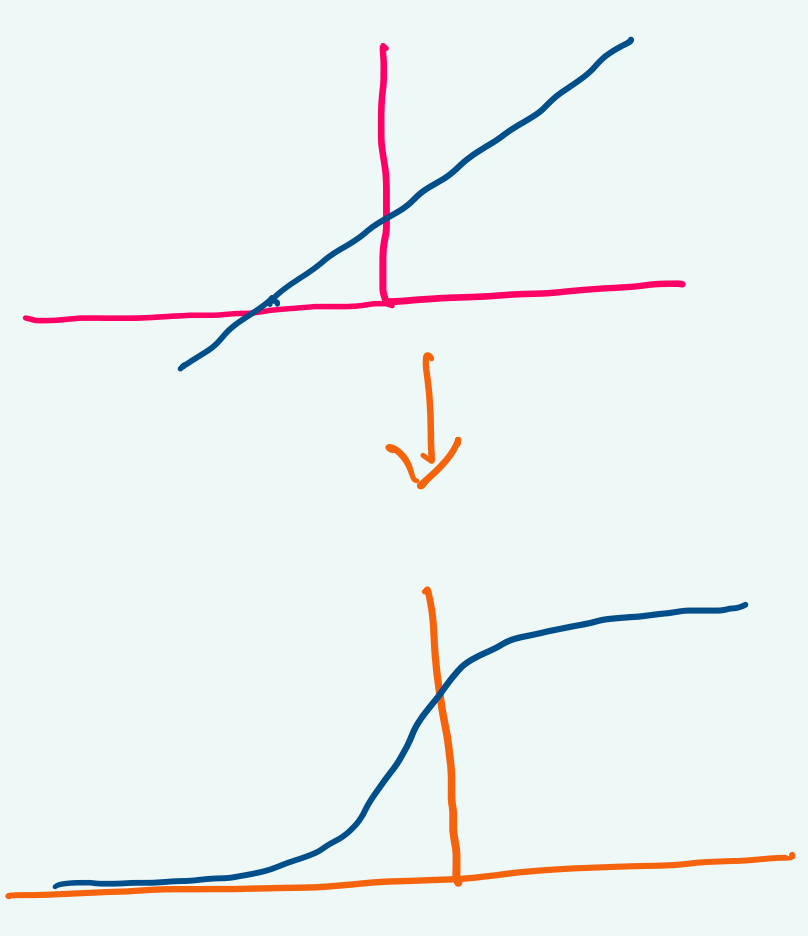
Metrics

Wednesday, 22 December 2021 7:55 PM

Hypothesis =

$\hat{y} = h_\theta(x) = \sigma(\theta^T x)$

$= \dfrac{1}{1+e^{-(\theta_0+\theta_1x_1+\theta_2x_2\ldots)}}$



Error / Loss / Cost

Binary Crossentropy / log-loss

$$L = -\frac{1}{m}\sum_{i=1}^{m} y^{(i)}\cdot\log\hat{y}^{(i)} + (1-y^{(i)})\log(1-\hat{y}^{(i)})$$

Gradient

$$\frac{\partial L}{\partial \theta_j} = \frac{1}{m}\sum_{i=1}^{m}(\hat{y}^{(i)} - y^{(i)})\cdot x_j^{(i)}$$

Gradient Descent

Repeat until converge {

$\theta_j = \theta_j - \alpha\cdot\dfrac{\partial L}{\partial \theta_j}$

}

$X = [\ \vdots\ \text{—x—}\ ]$

$\theta_0 x_0$, $x_0 = 1$

$\theta^T x \cdot \approx X\cdot\theta$ → $(m,n)\times(n,1)$ → $\sigma((m,1))$ → $= (m,1)$

$[y^{(1)}, y^{(2)}, y^{(i)}, \ldots, y^{(m)}] * [\log(\hat{y}^{(1)}), \log(\hat{y}^{(2)}), \ldots, \log(\hat{y}^{(m)})]$

np.mean([$y^{(1)}\log\hat{y}^{(1)}$ + — ; $y^{(i)}\log\hat{y}^{(i)}$ + — ; + —]) $(m,1)$

Hypothesis = $\sigma(\theta_0 + \theta_1x_1 + \theta_2x_2)$

$\theta_0 + \theta_1 x_1 + \theta_2 x_2 = 0$

$x_2 = \dfrac{-\theta_0 - \theta_1 x_1}{\theta_2}$

$(x_1^{(1)}, x_2^{(1)})$

$(x_1^{(2)}, x_2^{(2)})$

Regularization? → penalise coeff $[\theta_1, \theta_2, \ldots \theta_n]$

① overfitting / Underfit

Loss = log-loss + Regularization loss — L2

$$Loss = -\frac{1}{m}\sum_{i=1}^{m}\left[y^i\log\hat{y}^i + (1-y^i)\log(1-\hat{y}^i)\right] + \lambda\sum_{j=1}^{n}|\theta_j|^2$$

λ: regularization const, hyperparameter.

$\theta^T\theta = \theta^2$

$\|\theta_j\|_2$

$\theta_0 + \theta_1 x_1 + \ldots \theta_n x_2$ → 0

$$Loss = \text{log-loss} + \lambda\sum_{j=1}^{n}|\theta_j| \quad \rightarrow L1$$

→ create sparsity

L1 → 0

L2 ≠ 0 ≈0.0002

Elastic Net

$$Loss = \text{log-loss} + \underbrace{\lambda_1\sum_{j=1}^{n}|\theta_j|}_{L1} + \underbrace{\lambda_2\sum_{j=1}^{n}\theta_j^2}_{L2}$$

100000 × 2

200000

$\theta_1 = \theta_2 = \theta_3 = \ldots \theta_n \Rightarrow 0$

θ_0

$+\lambda_2\sum_{j=1}^{n}\theta_j$ gradient

C↓ λ↑ → underfit

C↑ λ↓ ≈0 → overfit

EVALUATION METRICS

① **Accuracy**

100 points → 60 correct, → 40 incorrect

$Acc = \frac{60}{100} \Rightarrow 0.6$

(M1) → 90% acc. (test). Is this model sensible?

Medical → Cancer (1000 people) class→1; → Not Cancer (9000 people) class→0

Imbalanced Dataset

input → (M0) → 0

[Dumb Model]

testset → 90 ✓, → 10 ✓

(90%)

(99%)

② **Confusion Matrix**

Binary classification (0 | 1) — Noncancer, Cancer; 0 → negative; 1 → +ve class

Matrix 2×2

3class Matrix 3×3

C class Matrix C×C

y: actual label; ŷ: Predicted labels

CM (actual labels vs predicted labels):

Predicted \ actual	0	1
0	a	b
1	c	d

a ⇒ actual=0 & predicted=0

b ⇒ actual=1 & predicted=0

c ⇒ actual=0 & predicted=1

d ⇒ actual=1 & predicted=1

Best Model ⇒ a↑ & d↑ (Right Prediction) | b & c↓ (wrong Prediction)

Predicted \ actual	0	1
0	TN	FN
1	FP	TP

T P

F P

T N — N ↳ what is your prediction? T ↳ are you correct?

N : # negative = TN+FP

P : # positive = FN+TP

① TPR = True Positive Rate = $\frac{TP}{P} = \frac{TP}{FN+TP}$

② TNR = True Negative Rate = $\frac{TN}{N} = \frac{TN}{TN+FP}$

③ FPR = false Positive Rate = $\frac{FP}{N} = \frac{FP}{TN+FP}$

④ FNR = $\frac{FN}{P}$

e.g → Imbalanced Test set → 900 -ve (Non cancer), → 100 +ve (cancer)

Predicted \ actual	0	1
0	850 (TN)	6 (FN)
1	50 (FP)	94 (TP)

TPR = 94% ⇑

TNR = $\frac{850}{900}$ ⇑

FPR = $\frac{50}{900}$ ⇓

FNR = 6% ⇓

} Sensible Model

Dumb Model

P \ A	0	1
0	900	100
1	0	0

TPR = 0%

TNR = 100%

FPR = 0%

FNR = 100%

} Dumb model

TPR ⇑ v.high

FNR ⇓ v.low ≈0% [critical]

FPR ≈4%.5%

Precision, recall, F1 score.

P \ A	0	1
0	TN	FN
1	FP	TP

$Pr = \frac{TP}{TP+FP}$

$Pr = \frac{\text{right}}{\text{all +ve prediction}}$

$Recall = TPR = \frac{TP}{TP+FN}$

$f1\ score = \frac{2PR}{P+R} \Rightarrow \frac{2}{\frac{1}{P}+\frac{1}{R}}$

recall = 0%, Pr = 0% → f1 Score = 0%

Best f1 score = 1 when Pr = 1, Recall = 1

$\frac{2\times1\times1}{1+1} \Rightarrow \frac{2}{2} = 1$

ROC-AUC