EVALUATION METRICS

09 1) Classification Accuracy

10 2) Logarithmic Loss

11 3) Confusion Matrix @100days-ofcode.

12 4) Area under curve

13 5) F1 Score

14 6) Mean Absolute Error

15 7) Mean Squared Error

16

(1) Classification Accuracy:-

17 Accuracy = No. of Correct predictions / Total No. of predictions made

18

(2) Log Loss – penalises false classifications.

Notes

$$\text{Logloss} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M} y_{ij} * \log(p_{ij})$$

16
THURSDAY

16.01.2014
Wk 03 . 016 - 349
JANUARY

JANUARY '14

WK	M	T	W	T	F	S	S
01	*	*	1	2	3	4	5
02	6	7	8	9	10	11	12
03	13	14	15	16	17	18	19
04	20	21	22	23	24	25	26
05	27	28	29	30	31	*	*
*	*	*	*	*	*	*	*

y_{ij} = indicates whether sample i belongs to class j or not

p_{ij} = indicates prob of sample i belonging to class j.

$[0, \infty)$ near 0 ⇒ high accuracy

Logloss ↓ Accuracy ↑

@100days_ofcode

③ Confusion Matrix :-

	Predicted No	Predicted Yes
Actual No		
Actual Yes		

Notes

④ AUC :-

True +ve Rate = True +ve / FN + TP

+ve data pts correctly represented as +ve

17 FRIDAY

17.01.2014
Wk 03 . 017 - 348
JANUARY

FEBRUARY '14

Wk	M	T	W	T	F	S	S
05	*	*	*	*	*	1	2
06	3	4	5	6	7	8	9
07	10	11	12	13	14	15	16
08	17	18	19	20	21	22	23
09	24	25	26	27	28	*	*
*	*	*	*	*	*	*	*

- FP Rate = $\frac{FP}{FP + TN}$

- we pts mistaken as +ve

09

10

⇒ [0,1] TP, FP Rate.

11

(5) F1 Score - tells how precise a classifier is. @100days-of-code

High precision, low recall

F1 Score ↑ be performance ↑

(6) Mean Absolute Error = $\frac{1}{N} \sum_{F1}^{N} |y_j - \hat{y}_j|$

17

18

LOGISTIC REGRESSION

classification Algo → Binary

→ Multi-Linear functions fails class

Notes

- Fraud or not
- Spam or not

⇒ Cost func = Signoid func.

$0 \leq h_\theta(x) \leq 1$

18 SATURDAY

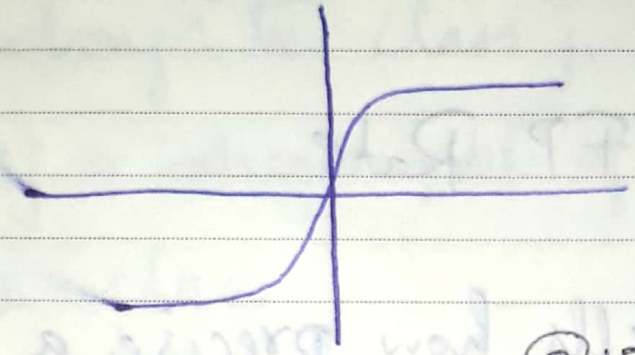
18.01.2014
Wk 03 . 018 - 347
JANUARY

JANUARY '14

Wk	M	T	W	T	F	S	S
01	*	*	1	2	3	4	5
02	6	7	8	9	10	11	12
03	13	14	15	16	17	18	19
04	20	21	22	23	24	25	26
05	27	28	29	30	31	*	*
*	*	*	*	*	*	*	*

09 Sigmoid func $\sigma_z = \frac{1}{1+e^{-z}}$

10



11 $z = \sum w_i x_i + \text{bias}$

12 @100days-of code

13 $\sigma(z) = \sigma(\beta_0 + \beta_1 x)$

14 One nominal var with 2 values

15 (Yes/No...) & 1 measurement var.

16 Nominal var = dependent var

17 Measurement var = Indep-

18 (1) Binary → dependent var is binary

(2) Multinomial → dependent var

19 Sunday nominal with more than 2 levels.