Last Class (August 12)

1) Recap — Quizzes
2) AT&T Churn Prediction Problem
3) Recap of Linear Regression
4) Intro to Logistic Regression
5) Thresholding and Step Function
6) Sigmoid Function
7) Geometric Interpretation
8) Maximum Likelihood
9) Logloss and optimization

Today's class

1) Overview of AT&T Churn Prediction
2) Accuracy Metric
3) Hyper-parameter Tuning
4) Logit / Log Odds
5) Impact of Outliers
6) Multi-class classification — OVR vs multi-nomial

100%

↳ 80% train + val

20% test

75% of 80 → 60% ← 75% train

25% of 80 → 20% 25% val

60 → train

20 → val

20 → test

bias (b)

$$z^{(i)} = w_0 + w_1 \cdot x_1^{(i)} + w_2 \cdot x_2^{(i)} + \dots + w_5 \cdot x_5^{(i)}$$

coeff_ → $[w_1, w_2, w_3, w_4, w_5]$

intercept_ → w_0

$$\text{sigmoid}(z^{(i)}) = \frac{1}{1 + e^{-z^{(i)}}}$$

↳ $\hat{y}^{(i)} = p(y = 1 / x^{(i)})$

output probability

Params for LogReg : weights and bias

Hyper-params : λ_1, λ_2, degree, η, num_epochs, threshold

I

clss 0 : 3000

clss 1 : 1500

II

clss 0 : 10,000

clss 1 : 1000

↓

Imbalanced

III

clss 0 : 2500

clss 1 : 2400

↳ Balanced

??

warm start :

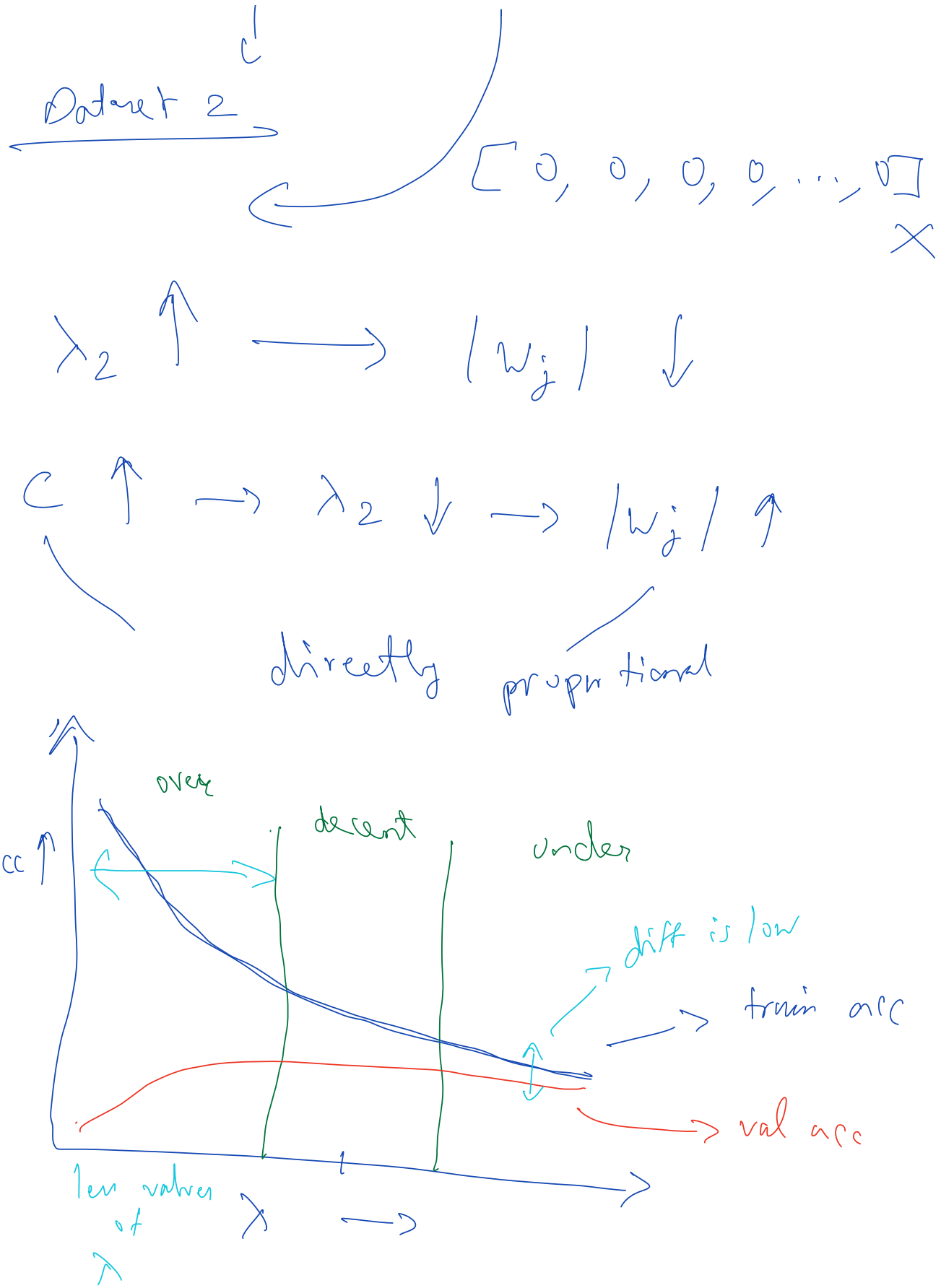
Dataset I

↓

LogReg

$w_0, w_1, w_2, \cdots w_d$

$[1.0, 2.0, -1.1, 2.05, \cdots]$


Dataset 2
[0, 0, 0, 0, ..., 0]
X
λ_2 ↑ → $|w_j|$ ↓
C ↑ → λ_2 ↓ → $|w_j|$ ↑
directly proportional
over
decent
under
Acc ↑
diff is low
train acc
val acc
ler values of λ
λ →

$$p = p(y=1 / x^{(i)}) = \hat{y}^{(i)}$$

$$\downarrow$$

$$\sigma(z^{(i)})$$

$$\swarrow$$

$$z^{(i)} = w^T x^{(i)} + w_0$$

$$p = \sigma(z)$$

$$= \frac{1}{1+e^{-z}} = \frac{e^z}{e^z+1}$$

$$p = \frac{e^z}{e^z+1} \Rightarrow \frac{e^z+1}{e^z} = \frac{1}{p}$$

$$\Rightarrow 1 + \frac{1}{e^z} = \frac{1}{p}$$

$$\Rightarrow \frac{1}{e^z} = \frac{1}{p} - 1 = \frac{1-p}{p}$$

$$\Rightarrow e^z = \frac{p}{1-p}$$

$$\Rightarrow \boxed{z = \log\left(\frac{p}{1-p}\right)}$$

p = prob. of success $\rightarrow$ prob. of belonging to class 1 (+ve class)

$1-p$ = prob. of failure → prob. of belonging to class 0 (−ve class)

$$\text{odds} = \frac{\text{prob. of success}}{\text{prob. of failure}}$$

$$= \frac{p}{1-p}$$

z = log of odds $= \log\left(\frac{p}{1-p}\right)$

$z \longleftrightarrow p$

→ inverse of each other

$p = \sigma(z) = \frac{1}{1+e^{-z}}$

$z = \log\left(\frac{p}{1-p}\right)$

$$\text{logit} = \log \text{ of odds}$$

$$= \log_e\left(\frac{p}{1-p}\right)$$

$$-\left[y \log(\hat{y}) + (1-y) \log(1-\hat{y})\right]$$

$y = 1, \quad 1 - y = 0$

$$-\left[1 \times \log(\hat{y})\right]$$

$$= -\log(\hat{y})$$

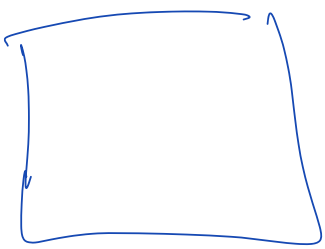
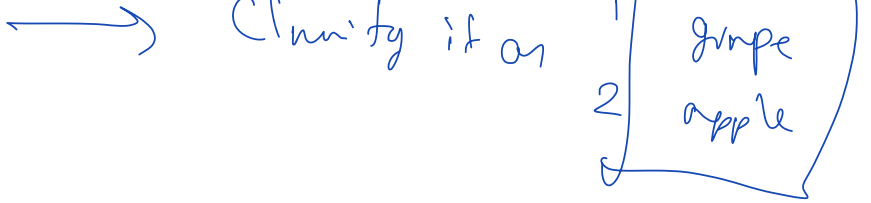
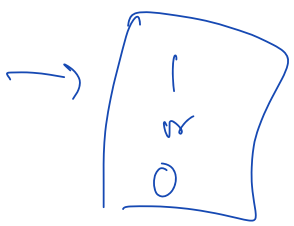


image of a fruit





OVR → one vs rest

or

one vs

Classifier 1

Data: apple, non-apple

apples vs non-apples

apple → (1) → +ve

grape, orange → non-apple → (0) → -ve

Classifier 2

grape vs non-grapes

grape → class 1 → +ve

orange, apple → class 0 → -ve

Classifier 3

orange, non-orange

orange → (1) → +ve

apple, grape → (0) → -ve

x_q →

Classifier Apple → $P_1(y = \text{apple} / x_q)$

Classifier Orange → $P_2(y = \text{orange} / x_q)$

Classifier grape → $P_3(y = \text{grape} / x_q)$

max (P_1, P_2, P_3)

↳ P_2 is max, → orange

P_1 is max $\Rightarrow$ o/p → apple

P_3 is max $\Rightarrow$ o/p → grape

P_1 P_2 P_3

0.1	0.01	0.05

clm 1

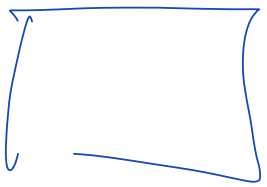


image of banana

0.9 0.8 0.85

clm 1

A O G

clm 1 clm 2 clm 3

0.1	0.08	0.12

clm 3

$$\sigma(z) = \frac{e^z}{1+e^z}$$

clm 1
clm 0

$x^{(i)}$

3 classes

$w_{01}, w_{11}, w_{21}, w_{31}, \ldots w_{d1}$ clm 1

$w_{02}, w_{12}, w_{22}, w_{32}, \ldots w_{d2}$ clm 2

$w_{03}, w_{13}, w_{23}, \ldots w_{d3}$ clm 3

$$x^{(i)} \longrightarrow x_1^{(i)}, x_2^{(i)}, x_3^{(i)}, \dots x_d^{(i)}$$

$$z_1^{(i)} = w_{01} + w_{11} \cdot x_1^{(i)} + w_{21} \cdot x_2^{(i)} + \dots w_{d1} \cdot x_d^{(i)}$$

$$z_2^{(i)} = w_{02} + w_{12} \cdot x_1^{(i)} + w_{22} \cdot x_2^{(i)} + \dots w_{d2} \cdot x_d^{(i)}$$

$$z_3^{(i)} = w_{03} + w_{13} \cdot x_1^{(i)} + w_{23} \cdot x_2^{(i)} + \dots w_{d3} \cdot x_d^{(i)}$$

$$P\left(y = \text{clan } 1 \,/\, x^{(i)}\right) = P_1 = \frac{e^{z_1^{(i)}}}{e^{z_1^{(i)}} + e^{z_2^{(i)}} + e^{z_3^{(i)}}} \text{ (Summation)}$$

$$P\left(y = \text{clan } 2 \,/\, x^{(i)}\right) = P_2 = \frac{e^{z_2^{(i)}}}{\text{Sum}\left(e^{z}\right) \rightarrow \sum_{k=1}^{3} e^{z_k^{(i)}}}$$

$$P_3$$

$$P_1 + P_2 + P_3 = 1$$

$$P_1 = \frac{e^{z_1^{(i)}}}{\Sigma}, \quad P_2 = \frac{e^{z_2^{(i)}}}{\Sigma}, \quad P_3 = \frac{e^{z_3^{(i)}}}{\Sigma}$$

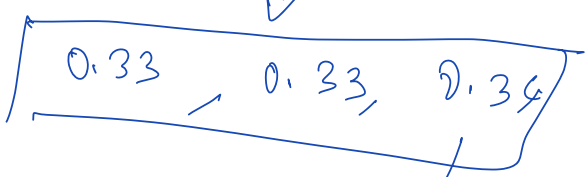
$$P_1 + P_2 + P_3 = \frac{e^{z_1} + e^{z_2} + e^{z_3}}{\Sigma} = 1$$

Multi-nomial

0.1 , 0.08 , 0.12 ✗

sum (prob) = 1

image of banana → multinomial log reg



clus 3

↓

grape

max prob > 0.5

↳ classify

else say un-recognized

$$\log\left(\frac{p}{1-p}\right) < 0$$

$$\log\left(\frac{p}{1-p}\right) < \log(1)$$

$$\frac{p}{1-p} < 1 \Rightarrow p < 1-p$$

$$\Rightarrow 1-p > p$$

$$\log\left(\frac{p}{1-p}\right) = z \rightarrow \log(\text{odds})$$

$$p = \sigma(z)$$

$\log(\text{odds}) = z$

$z = w^T x + b$

z & x —, linear