Lecture - Logistic Reg-3

https://colab.research.google.com/drive/1m1OhqfsbOBP8jVtOaV1DQg_qTczuoND5?usp=sharing

Agenda

- GD for Logistic Regression Loss = MSE Lin Reg / NLL Log Reg
- Sklearn Implementation
- Evaluation Metric for Classification R^2 Lin Reg
- Logodds
- Multiclass Classification
- Impact of Outliers OPTIONAL

Revision

$$\hat{y} = \sigma(\underbrace{w^T x + w_0}_{\text{same as linear Reg } z})$$

$\hat{y} \rightarrow p$

$\sigma \rightarrow$ sigmoid

$$= \frac{1}{1+e^{-z}} \quad \text{where } z = w^T x + w_0.$$

Loss function for Log Reg

1. $p,\ 1-p \longrightarrow (p^i)^{y^i} \cdot (1-p^i)^{1-y^i}$
2. Independent Event & m samples $\prod$ — ①
3. Log both sides — ②
4. Negative of this — ③

) Max. Likelihood estimation (MLE)

NLL = Neg. Log Likelihood.
3 2 1

Optimisation:

Gradient Descent

1/0

$$w_j \rightarrow w_j - \alpha \frac{\partial L}{\partial w_j}$$

Neg. log Likelihood

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

Derivative (Post-Read)

$$\frac{\partial \log x}{\partial x} = \frac{1}{x} \quad \Big| \quad \frac{\partial \sigma(z)}{\partial z} = \sigma(z)(1-\sigma(z)) \quad \Big| \quad \frac{\partial z}{\partial w} = x$$

Post Read.

w^Tx

$$\frac{\partial NLL}{\partial W} = -\sum_{i=1}^{m} \underbrace{y^i \log \hat{y}^i}_{\mathrm{I}} + \underbrace{(1-y^i)\log(1-\hat{y}^i)}_{\mathrm{II}}$$

I

$$\frac{\partial (y \log \hat{y})}{\partial W} = y\frac{\partial(\log \hat{y})}{\partial W} = y \cdot \frac{\partial \log \hat{y}}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial W}$$

CHAIN RULE

$$= y \cdot \frac{1}{\hat{y}} \cdot \frac{\partial \hat{y}}{\partial W}$$

$\sigma(z)$

CHAIN RULE

$$= \frac{y}{\hat{y}} \cdot \frac{\partial \sigma(z)}{\partial z} \cdot \frac{\partial z}{\partial W}$$

$W^T x + w_0$

$$= \frac{y}{\hat{y}}\, \sigma(z)(1-\sigma(z)) \cdot \frac{\partial z}{\partial W}$$

$\hat{y}$

$$= \frac{y}{\hat{y}} \cdot \sigma(z)(1-\sigma(z)) \cdot X$$

$$= \frac{y}{\cancel{\hat{y}}} \cdot \cancel{\hat{y}}(1-\hat{y}) \cdot X$$

$$\boxed{\mathrm{I} = y(1-\hat{y}) \cdot X}$$

$$\mathrm{II} \quad \frac{\partial (1-y)\log(1-\hat{y})}{\partial W} = (1-y)\frac{\partial \log(1-\hat{y})}{\partial W}$$

$$=$$

$\vdots$

Derivative of 1st term = $y(1-\hat{y})\cdot x$

Derivative of 2nd term = $-(1-y)\,\hat{y}\cdot x$

$$\frac{\partial NLL}{\partial w} = -\left[y(1-\hat{y})x - (1-y)\cdot\hat{y}\,x\right]$$

$$= -x\left[y(1-\hat{y}) - (1-y)\cdot\hat{y}\right]$$

$$= -x\left[y - \cancel{y\hat{y}} - \hat{y} + \cancel{y\hat{y}}\right]$$

$$= -x\cdot\underline{(y-\hat{y})} \text{ — error (e)}$$

$$\frac{\partial L}{\partial w} = -e\cdot x \text{ — LogReg}$$

LinReg

$$\frac{\partial L}{\partial w} = -2\cdot e\cdot x$$

$\downarrow$

constant

Gradient Descent equation for Lin Reg and Log Reg are same.

$$\boxed{w_j \rightarrow w_j - \alpha \sum^{m} 2 \cdot (y - \hat{y}) \cdot x}$$

Generalized Linear Models

	LinReg	LogReg
Loss	MSE	NLL
$\frac{\partial L}{\partial w}$	$\sum -2 \cdot e \cdot x$	$\sum - e \cdot x$

sklearn - Colab Notebook

Evaluation Metric for Classification

~~R2/AdjR²~~

y^i	$\hat{y}^i$	
1	1	✓
0	0	✓
1	0	×
0	0	✓
0	1	×

$$\text{Accuracy} = \frac{\#\ \text{correct prediction}}{\#\ \text{total pred}}$$

$$= \frac{3}{5}$$

$$= 60\%$$

Log-odds interpretation for LogReg

Odds of success

Odds of a horse winning a game.

Success : Failure

4 : 1

success to failure ratio

$$P(\text{winning}) = \frac{4}{4+1} = \frac{4}{5}$$

$$P(\text{failing}) = \frac{1}{4+1} = \frac{1}{5}$$

$$\text{Odds} = \frac{P(\text{winning})}{P(\text{Failure})} = \frac{4/\cancel{5}}{1/\cancel{5}} = 4:1$$

Success in LogReg $P(y=1|x)$

$P(W) = p$

$P(L) = 1-p$

$$\boxed{\text{Odds} = \frac{p}{1-p}}$$

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

lly $$1-p = 1 - \frac{e^z}{e^z+1} = \frac{\cancel{e^z}+1-\cancel{e^z}}{e^z+1} = \frac{1}{e^z+1} \quad 1-p$$

$$\text{Odds} = \frac{p}{1-p} = \frac{e^z/\cancel{e^z+1}}{1/\cancel{e^z+1}} = e^z$$

$$\text{Odds} = e^z$$

$$\log_e \text{Odds} = \log_e e^z$$

$$\log \text{Odds} = z \cancel{\log_e e} \rightarrow 1$$

★★★

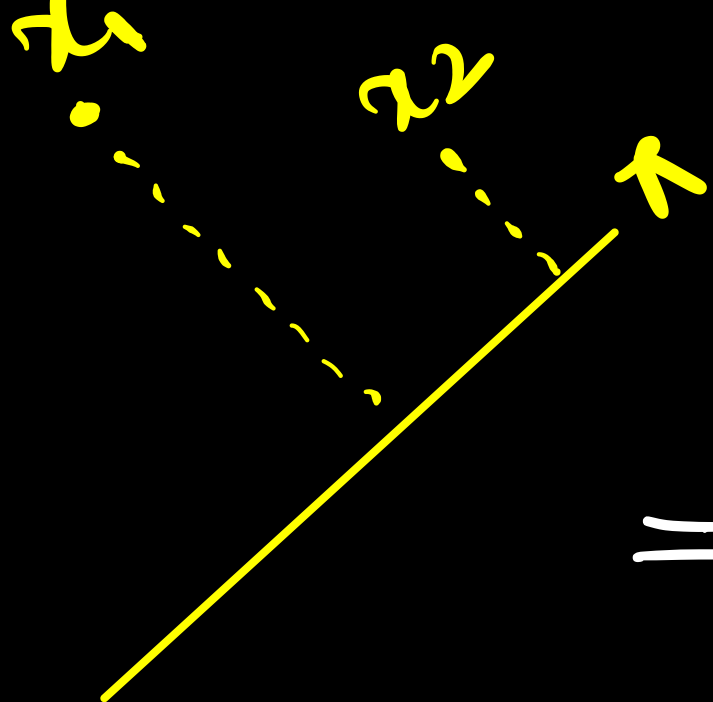
$$\boxed{\text{log-odds} = z}$$

Odds of occurring class-1

$w^T x + w_0$

$\propto$

Distance of x from a point



$d_{x_1} > d_{x_2}$ from π

log odds x_1 > log odds x_2

$\Rightarrow$ chances of x_1 belonging to class 1 are greater than chances of x_2 belonging x_2

larger distance = larger log-odds

Interview

LinReg

$$w^T x + w_0 = \hat{y}$$

LogReg

$$w^T x + w_0 = \log\left(\frac{p}{1-p}\right)$$

Multi-class classification using LogReg

LogReg - used for BINARY classification.

m=3, Orange, Apple, Grapes.

$$\begin{array}{c|ccc} O = & 0 & 0 & 1 \\ A = & 0 & 1 & 0 \\ G = & 1 & 0 & 0 \\ \hline & y_{M3} & y_{M2} & y_{M1} \end{array}$$

$$y \text{ shape} = (m, 3)$$

classes

samples

Solution?? MC task with Log R

Train 3 models.

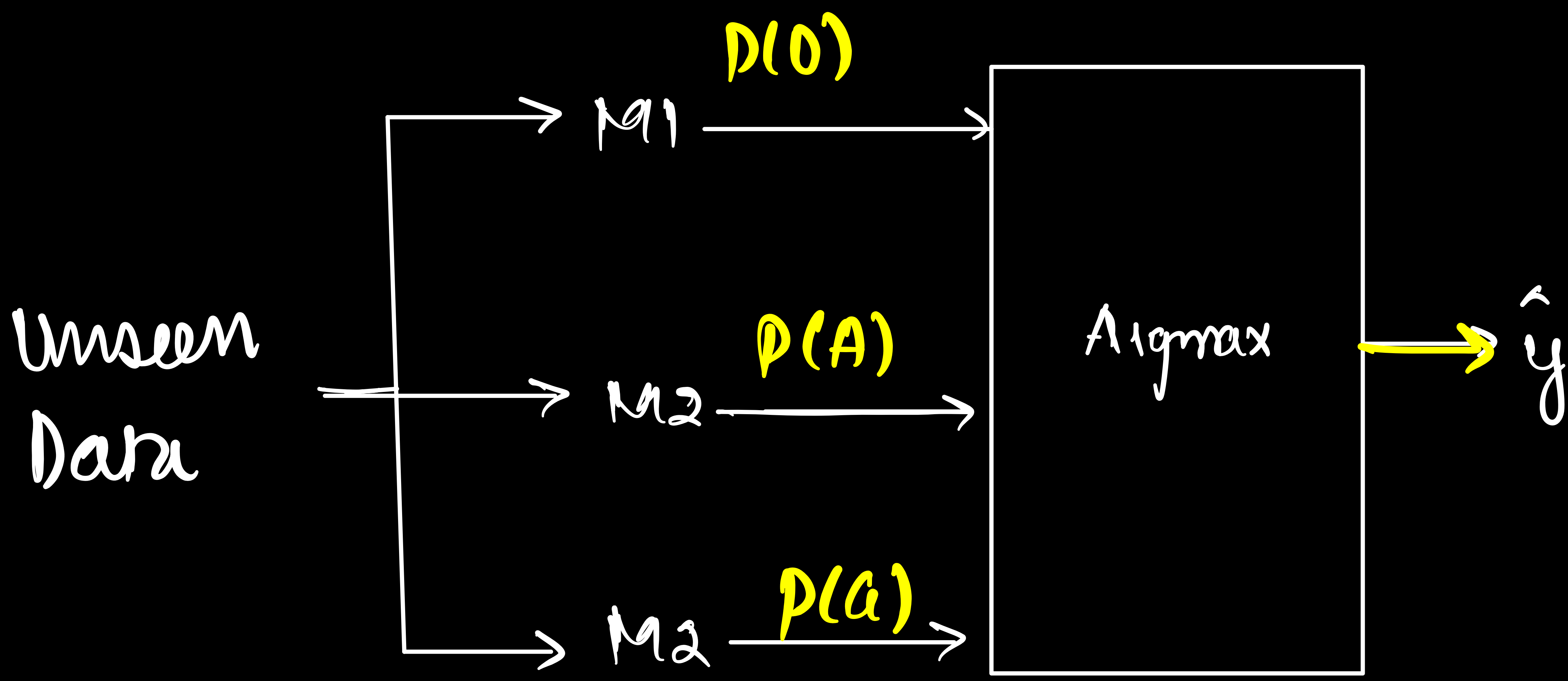
M1 : Orange or Not → p(Orange) p_1
1 0

M2 : Apple or Not → p(Apple) p_2
1 0

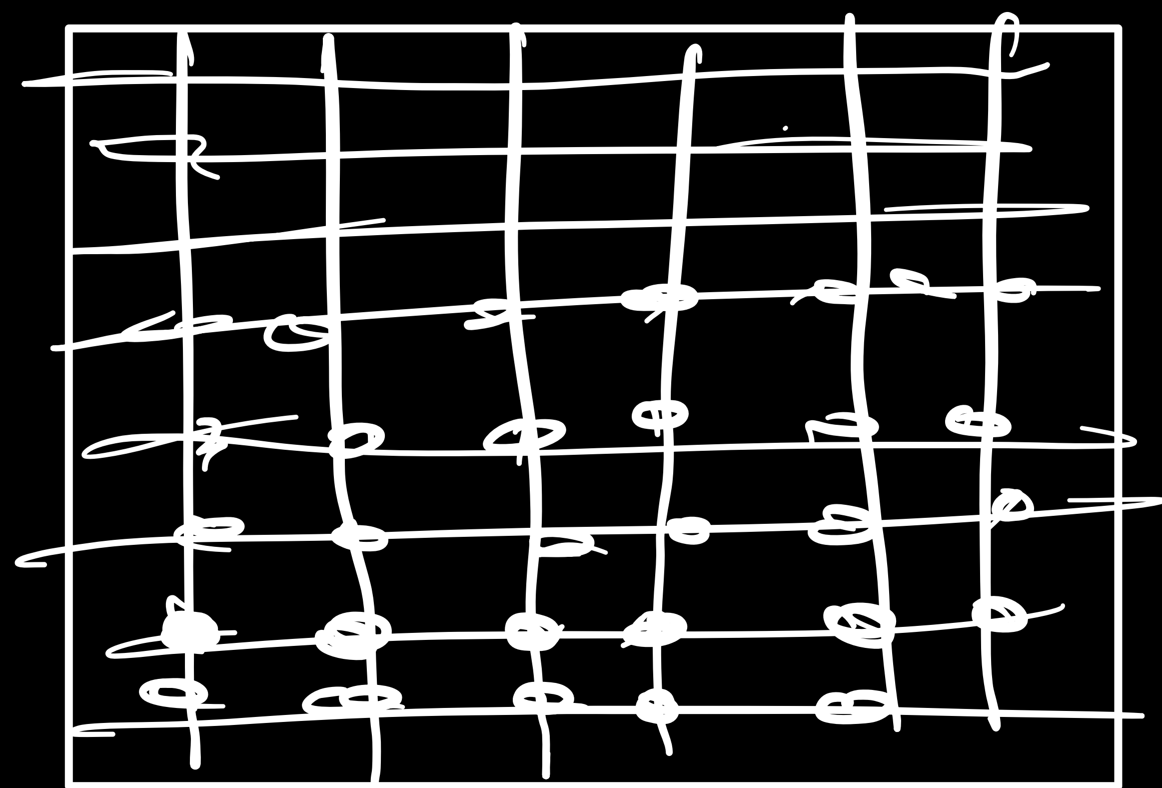
M3 : Grape or Not → p(Grape) p_3
1 0

$\text{argmax}(p_1 p_2 p_3)$ = FINAL LABEL

One-V/S-Rest classification-



lecture on
Fault in this approach- Neural Network

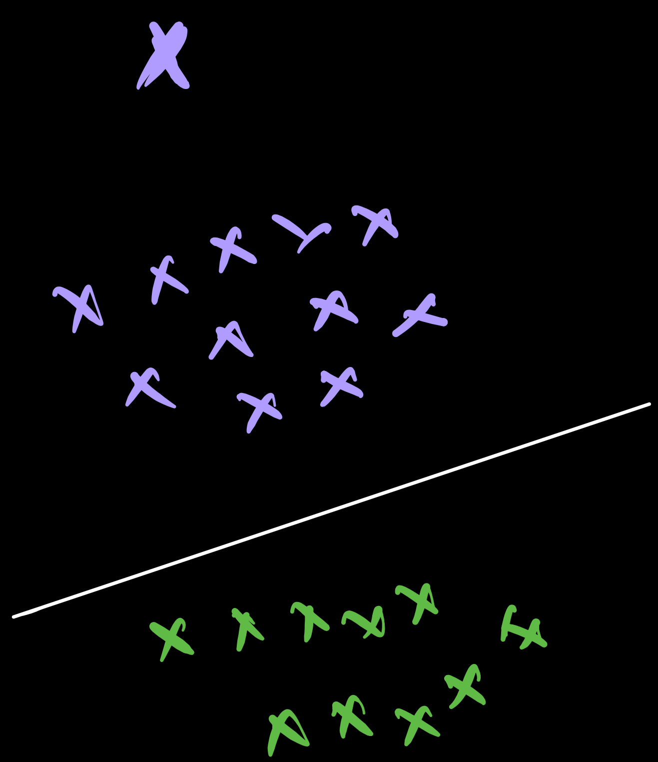


✓ np.meshgrid.

Impact of outliers

case1) Outlier is correctly predicted

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



case 2) Outlier is wrongly predicted