Eric Han
eric_han@nus.edu.sg
https://eric-han.com
Computer Science

T05 – 10 Oct 2024

Week 8

CS2109s TG35,36

Content I

Section 1: **Linear vs Non-linear Separability**

Decide whether a bunny is ready to be released into the wild based on two features: **Feature A** is a bunny's cuteness score and **Feature B** is a bunny's fluffiness score.

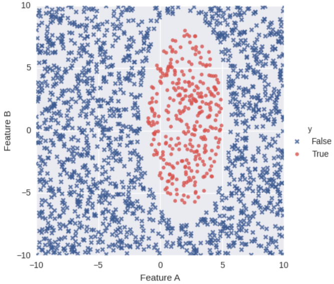


Figure 1: Feature A/B; Ready to be released into the wild?

1. Which *min* set of features that will perfectly (linearly) classify?

2 After changing production methods, samples are collected below; *min* features?

3 [@] How can we always find a *min* set of features, how does it relate to kernels?

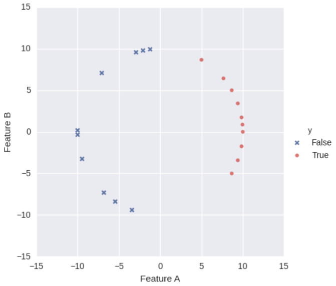


Figure 2: New Production Method.

Recap

- What is a transformation?
- What is linear separability, why is it desirable?
- How to achieve linear separability?

Notice that an ellipse with major and minor axis parallel to y-axis and x-axis is sufficient to classify the data. Hence,

- (A^2, B^2, A, B) minimally suffices.

For more general ellipses (or conics) you can use the more general set of features:

- (A^2, AB, B^2, A, B).

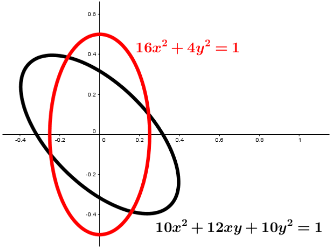


Figure 3: Centered Ellipse; If axis-parallel AB is not needed. If centered, A, B is not needed.

We can use just use A.

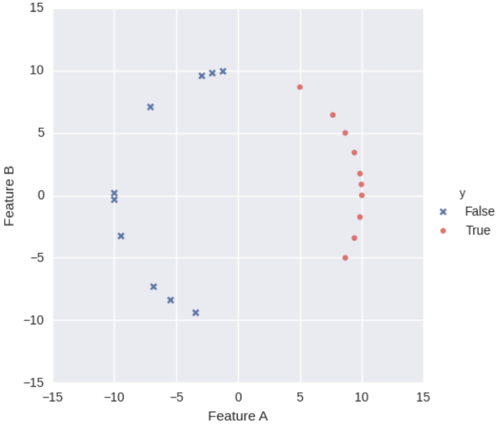
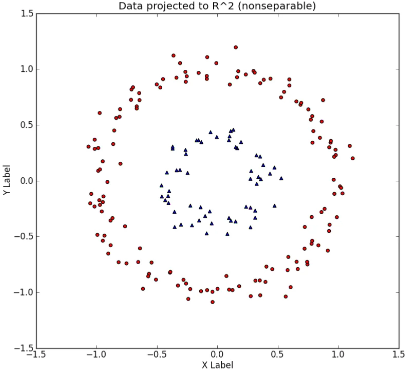


Figure 4: New Production Method.

Transformation for Linear Separability



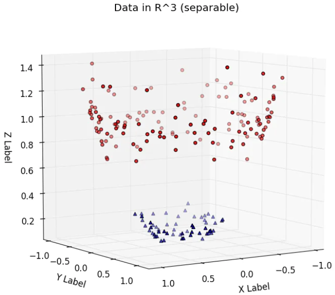


Figure 5: Illustration of transformation - Linear Separability

Section 2: **Loss Function of Logistic Regression**

Question [G]

Logistic Regression model which has the following hypothesis, where, $h_w(x)$ could be interpreted as a probability p assigned by the model such that $y = 1$. The probability of $y = 0$ is therefore $1 - p$.

$$h_w(x) = \frac{1}{1 + e^{-w^T x}}$$

1. Calculate the derivative of $\log(p)$ with respect to each weight w_i.
2. Calculate the derivative of $\log(1 - p)$ with respect to each weight w_i.
3. Derive $\frac{\partial L}{\partial w_i}$, where L is the loss function of logistic regression model.

Recap

- Why do we want to find $\frac{\partial L}{\partial w_i}$?
- What is logistic regression?
  - What is logistic? what is regression?

Answer 1

First we write the probabilty p as a function of x. $p = \frac{1}{1+e^{-w^T x}} = \frac{1}{1+e^{-w \cdot x}} = \frac{1}{1+e^{\sum_{i=1}^{n} -w_i x_i}}$

Take the log of both sides,

$\log(p) = \log\left(\frac{1}{1+e^{\sum_{i=1}^{n} -w_i x_i}}\right) = -\log(1 + e^{\sum_{i=1}^{n} -w_i x_i})$

Now we differentiate $\log(p)$ with respect to w_i

$$
\begin{aligned}
\frac{\partial \log(p)}{\partial w_i} &= -\left(\frac{1}{1+e^{\sum_{i=1}^{n} -w_i x_i}} \frac{\partial}{\partial w_i}(1 + e^{\sum_{i=1}^{n} -w_i x_i})\right) \\
&= -p \frac{\partial}{\partial w_i}(1 + e^{\sum_{i=1}^{n} -w_i x_i}) \\
&= -p(-x_i) e^{\sum_{i=1}^{n} -w_i x_i} \\
&= (1-p) x_i
\end{aligned}
$$

First we write the probabilty $1-p$ as a function of x.

$1-p = 1 - \frac{1}{1+e^{-w^T x}} = \frac{e^{-w^T x}}{1+e^{-w^T x}} = \frac{1}{1+e^{w^T x}} = \frac{1}{1+e^{w \cdot x}} = \frac{1}{1+e^{\sum_{i=1}^n w_i x_i}}$

Take the log of both sides,

$$\log(1-p) = \log\left(\frac{1}{1+e^{\sum_{i=1}^n w_i x_i}}\right) = -\log(1+e^{\sum_{i=1}^n w_i x_i})$$

Now we differentiate $\log(1-p)$ with respect to w_i

$$\begin{aligned}
\frac{\partial \log(1-p)}{\partial w_i} &= -\left(\frac{1}{1+e^{\sum_{i=1}^n w_i x_i}} \frac{\partial}{\partial w_i}(1+e^{\sum_{i=1}^n w_i x_i})\right) \\
&= -(1-p)\frac{\partial}{\partial w_i}(1+e^{\sum_{i=1}^n w_i x_i}) \\
&= -(1-p)(x_i)e^{\sum_{i=1}^n w_i x_i} \\
&= -(1-p)(x_i)\left(\frac{p}{1-p}\right) = -p x_i
\end{aligned}$$

Answer 3

$$L = -y\log(h_w(x)) - (1-y)\log(1-h_w(x))$$

First we substitute $h_w(x)$ as p:

$$L = -y\log(p) - (1-y)\log(1-p)$$

Now we differentiate L with respect to w_i:

$$\begin{aligned}
\frac{\partial L}{\partial w_i} &= -y\frac{\partial \log(p)}{\partial w_i} - (1-y)\frac{\partial \log(1-p)}{\partial w_i} \\
&= -y(1-p)x_i - (1-y)(-px_i) \\
&= -x_i(y-p) \\
&= x_i(h_w(x)-y)
\end{aligned}$$

Section 3: **Precision, recall, F1 score and ROC curve**

Question 1-3

Model M outputs 1 if $M(x)$ is greater than or equal to the threshold p, otherwise 0.

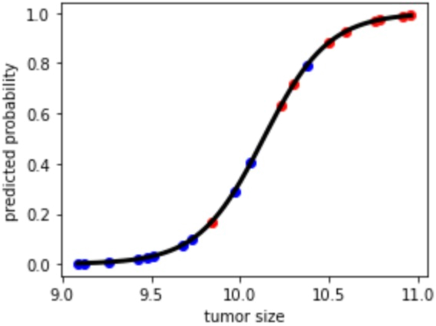


Figure 6: Model probability output and tumor size

1. For the threshold $p = 0.5$, come up with the confusion matrix.
2. For the threshold $p = 0.5$, find the precision, recall and F1 score.
3. Based on the figure, derive the ROC curve.

Answer 1

.	Prediction 0	Prediction 1
Actual 0	10	1
Actual 1	1	8

Answer 2

$$Precision = \frac{TP}{TP+FP} = \frac{8}{8+1} = \frac{8}{9}, Recall = \frac{TP}{TP+FN} = \frac{8}{8+1} = \frac{8}{9}$$

$$F1 = \frac{2 \times TP}{2 \times TP + FP + FN} = \frac{2 \times 8}{2 \times 8 + 1 + 1} = \frac{8}{9}$$

Answer 3

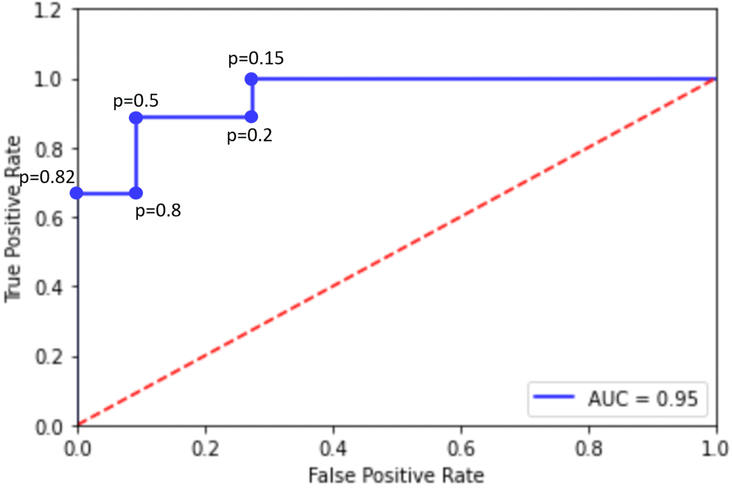


Figure 7: ROC curve

Question 4-7 [G]

4. Based on the ROC curve you derived, decide which threshold you want to choose among $p = 0.2$, $p = 0.5$ and $p = 0.8$.

[@] When to maximize precision or recall? What does it mean?

5. Detecting tumours
6. Detect plagiarism
7. Credit Card Fraud

Maximize precision / recall = Minimize FP / FN = Minimize Type 1 / Type 2 Error.

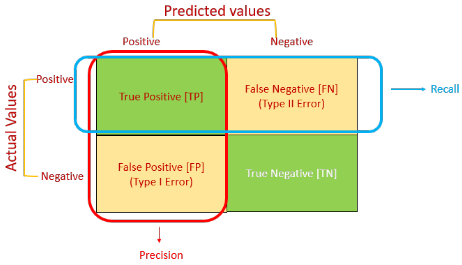


Figure 8: Intutition

For the application, which is more severe?

- Type 2 error - Missing diagnosis of tumor when actually tumor
- Type 1 error - Wrongly diagnosis of tumor when no tumor

If regular check up > Min start treatment on healthy > Min Type 1 > Max Precision

If monitoring > Min stop cancer treatment on sick > Min Type 2 > Max Recall

Section 4: **Logistic Regression for Multi-Class Classification**

Logistic Regression for Multi-Class Classification:

$$W = \begin{pmatrix} w_{cat} \\ w_{horse} \\ w_{elephant} \end{pmatrix} = \begin{pmatrix} 4.2 & -0.01 & -0.12 \\ -20 & -0.08 & 35 \\ -1250 & 0.82 & 0.9 \end{pmatrix}, \quad X = \begin{pmatrix} 1 & 4.2 & 0.4 \\ 1 & 720 & 2.4 \\ 1 & 2350 & 5.5 \end{pmatrix}$$

1. Compute the probability of an animal belonging to a certain class and classify them.
2. What if we want to extend the classification task to classify other animals? Can we train a new model while keeping the weights of the previous models?

Recap

1. What is the equation for Logistic Regression?
2. How can we compute this efficiently?

Answer 1

$$-X \times W^T = \begin{pmatrix} -4.1100 & 6.3360 & 1246.1960 \\ 3.2880 & -6.4000 & 657.4400 \\ 19.9600 & 15.5000 & -681.9500 \end{pmatrix}, P = \begin{pmatrix} 0.9839 & 0.0018 & 0.0000 \\ 0.0360 & 0.9983 & 0.0000 \\ 0.0000 & 0.0000 & 1.0000 \end{pmatrix}$$

$$Y = \begin{pmatrix} cat \\ horse \\ elephant \end{pmatrix}$$

Answer 2

If the new class has distinct features then yes. Otherwise no. However, the model may still benefit from retraining.

Section 5: **Evaluating Logistic Regression**

Question 5 [G]

Which of the following evaluation metrics is the **least** appropriate when comparing a logistic regression model's output with the target label?

a. Accuracy
b. Binary Cross Entropy Loss
c. Mean Squared Error
d. AUC-ROC
e. Mean Absolute Error (Added)

[@] What is the difference between evaluation metrics vs cost functions / loss? Which would be the best for LR loss?

Recap

1. Which methods are primarily used for classification?
2. What are some of the key limitations of each method?

Answer 5

Metrics	Type	Formula
Accuracy	Class	$\frac{TP+TN}{TP+FP+FN+TN}$
Binary Cross Entropy	Class Loss	$-y\log(h_w(x)) - (1-y)\log(1-h_w(x))$
Mean Squared Error	Reg. Loss	$\frac{1}{2}(y - h_w(x))^2$
Mean Absolute Error	Reg. Loss	$\frac{1}{2}(\|y - h_w(x)\|)$
AUC-ROC	Class	Area under a ROC curve

Abuse: Eg1 is better than Eg2 $y = [0, 0, 1], \hat{y}_1 = [0.4, 0.4, 0.6], \hat{y}_2 = [0.1, 0.6, 0.9]$, but

.	MSE	MAE	BCE
Eg1	0.08	0.20	0.511
Eg2	0.063	0.133	0.376

Depends on the task / objective (performance/model uncertainty) and context:

- Accuracy:
  - Dataset must be close to being uniform to be meaningful
- Binary Cross Entropy Loss:
  - Suffers from problem with being objective performance measure
  - Maybe appropriate if objective is model uncertainty comparing within LR classes
  - Designed for loss, popular and has properties to rely on:
    - Measure difference in 2 probability distribution
- MAE/MSE:
  - Suffers from problem with being objective performance measure
  - Designed for regression, essentially distance measures
- AUC-ROC:
  - Usually the most robust
  - More complicated to calculate

To help you further your understanding, not compulsory; Work for Snack/EXP!

Tasks

1. Implement code to solve C2,D1, no boilerplate code given.
   a. Calculation of precision, recall and F1 score for qn in section Precision, recall, F1 score and ROC curve.
   b. Calculation of probability and class for qn in section Logistic Regression for Multi-Class Classification.

Buddy Attendance Taking

1. [@] and Bonus declaration is to be done here; You should show bonus to Eric.
2. Attempted tutorial should come with proof (sketches, workings etc...)
3. Random checks may be conducted.
4. Guest student should come and inform me.

Figure 9: Buddy Attendance: https://forms.gle/q5Secb3dHshmXNXd7

References I