CS2109s - Tutorial 5

Eric Han (TG12-TG15)

Mar 14, 2024

Annoucements

Important admin

1. Congratulations on clearing your Midterm exams!
    1.1 When Midterm results? - Please be patient; Rizki/Chonghui is marking the scripts.
    1.2 When released any issues also go to them.
2. Last tutorial no [@] Questions, this week will have 4 [@] questions!

Question 1 [G]

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.

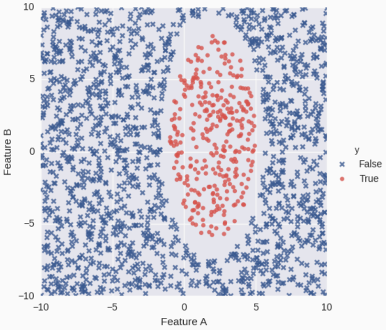

Figure 1: Feature A/B; Ready to be released into the wild?

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

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

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

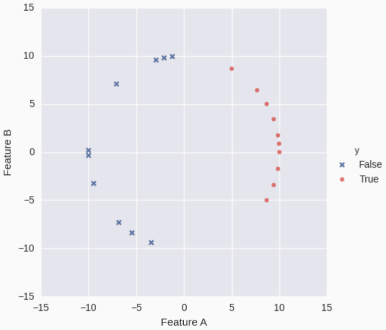

Figure 2: New Production Method.

Recap

- What is linear separability, why is it desriable?
- How to achieve linear separability?

Answer 1a

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).

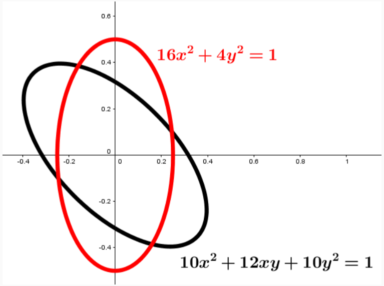

Figure 3: Centered Ellipse; If axis-parallel AB is not needed. If centered, A, B is not needed.

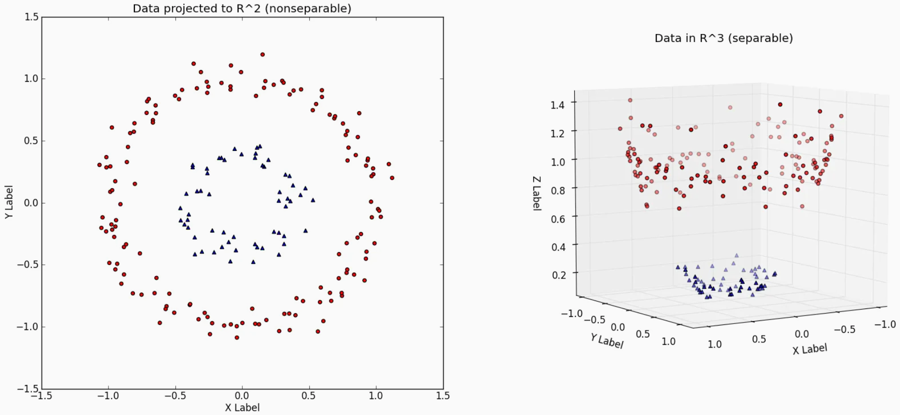

Figure 4: Linear Separability

Answer 1b

We can use just use A.

Question 2 [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 $1 - p$.

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

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

Recap

- What is logistic regression?
    - What is logistic? what is regression?

Answer 2a

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

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

Answer 2b

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 2c

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

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:

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

Question 3

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

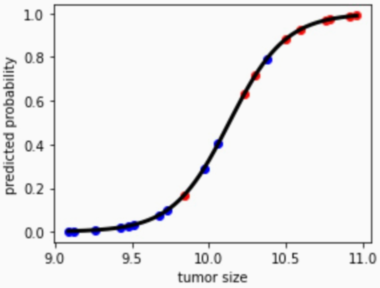

Figure 5: Model probability output and tumor size

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

Answer 3a

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

Answer 3b

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

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

Answer 3c

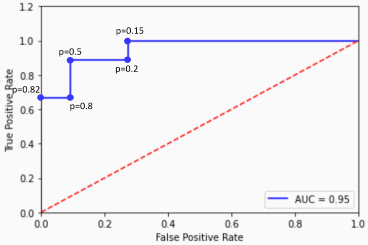

Figure 6: ROC curve

Question 3d-f [G]

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

e. [@] When to maximize precision or recall? What does it mean?
   - Detecting tumours / Detect plagiarism / PS5 Credit Card Fraud

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

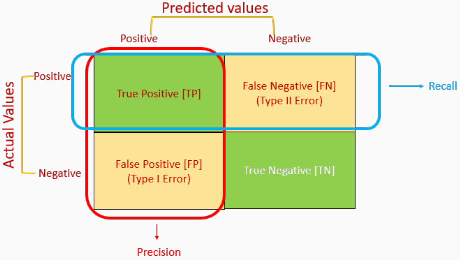

Figure 7: Intutition

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

Question 4

Logistic Regression for Multi-Class Classification:

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

a. Compute the probability of an animal belonging to a certain class and classify them.
b. 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 4a

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

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

Answer 4b

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

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

[@] 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	$\frac{1}{2}(\lvert y - h_w(x)\rvert)$
AUC-ROC	Class	Area under a ROC curve

Abuse: M1 is better than M2 $y = [0, 0, 1], \hat{y}_1 = [0.2, 0.4, 0.6], \hat{y}_2 = [0.1, 0.6, 0.9]$, but

.	MSE	MAE	BCE
M1	0.08	0.20	0.511
M2	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

Bonus Qn

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

Tasks

1. Implement code to solve Q3,4,5, no boilerplate code given.
    1.1 Calculation for Q3a,b precision and recall.
    1.2 Calculation for Q4a using `numpy` matrices
    1.3 Calculation for the Q5 illustration between MSE/MAE/BCE

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. Guest students must inform Eric and also register the attendance.

Figure 8: Buddy Attendance: https://forms.gle/jsGfFyfo9PTgWxib6