學號： r12227118　　系級： 心理所碩二　姓名： 張瑋宸

1. (1%) AUC

   a. (1%) 證明$AUC = Pr\{f(x^+) > f(x^-)\}$,
   其中x^+代表實際上 positive 的一個樣本, x^-代表實際 negative 的一個樣本, f表示模型預測該樣本 positive 的機率或程度。

Let $f(x^+) = X \sim g(\cdot), f(x^-) = Y \sim h(\cdot)$, and $G(\cdot), H(\cdot)$ are the cdf respectively. Moreover, assume $g(\cdot), h(\cdot)$ are independent, that is, the positive score a classifier gives to positive samples is independent of the score one gives to negative samples.

In ROC, AUC is the integral of True Positive Rate $(1-G(t))$ over False Positive Rate $(1-H(t))$, hence,

$$
\begin{aligned}
AUC &= \int_0^1 (1-G(t))d(1-H(t)) \\
&= \int_{-\infty}^{\infty} (1-G(t))h(t)dt \\
&= \int_{-\infty}^{\infty} P(X>t)P(Y=t)dt \\
&= \int_{-\infty}^{\infty} P(X>t, Y=t)dt \\
&= P(X-Y>0) = P\left(f(x^+) > f(x^-)\right)
\end{aligned}
$$

2. (1%) Surrogate Loss

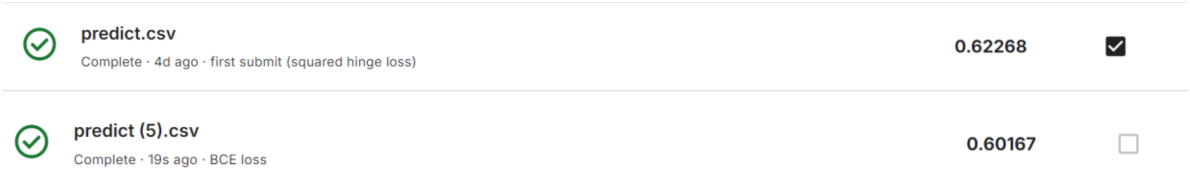
   a. (0.5%) 實作一個 surrogate loss 並貼上對應的程式碼。
   b. (0.5%) 請簡短介紹你使用的 loss, 並比較使用 BCELoss 跟 Surrogate Loss 在 Public Leaderboard 上的表現。

a.

```
# assited by gemini
def loss_function(y_pred, y_true):

   positive_mask = y_true == 1
   negative_mask = y_true == 0

   positive_predictions = y_pred[positive_mask]
   negative_predictions = y_pred[negative_mask]

   # Broadcasting to avoid nested loops
   diff = positive_predictions.unsqueeze(1) - negative_predictions  # Shape: (num_pos, num_neg)
   loss = torch.sum((torch.clamp(1 - diff, min=0)**2))  # maximum: squared hinge loss
 return loss
```

b. 使用作業投影片提供之 Squared Hinge Loss:

- Squared Hinge Loss $\mathcal{L}_{\text{squared hinge}} = \sum_{x^+\in\mathcal{P}} \sum_{x^-\in\mathcal{N}} \max(0, 1-(f(x^+)-f(x^-)))^2$



BCE loss 比起 Squared Hinge Loss 分數稍微低一點。但我最後採用的模型是使用 XGBoost，採用內建的 loss function 仍是 BCE loss ('objective': 'binary:logistic')

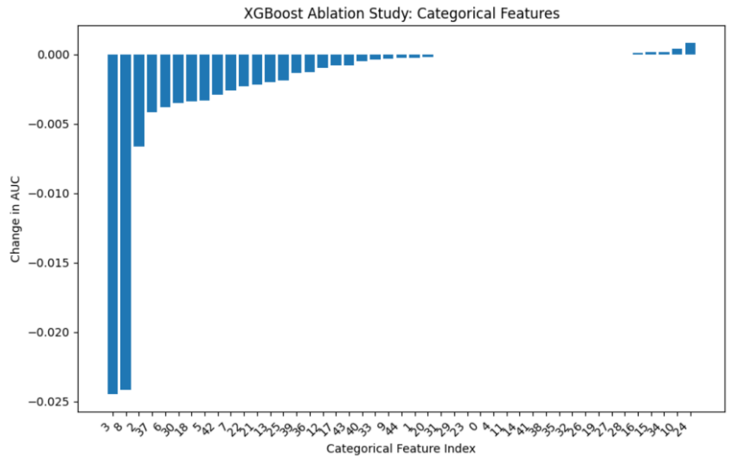
3. (1%) Feature importance

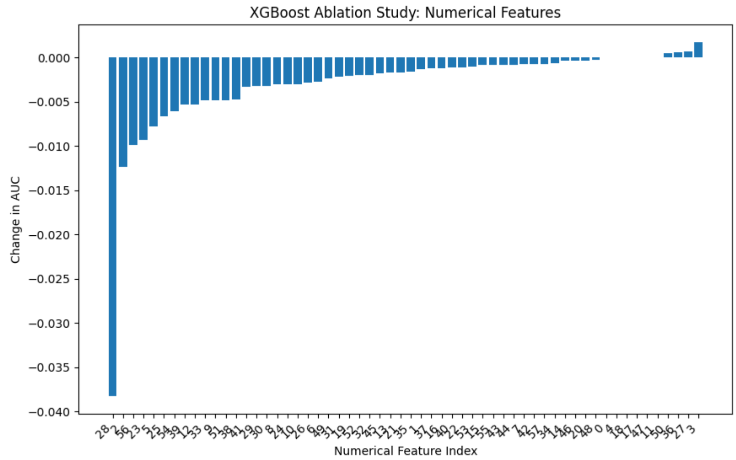
   a. 使用 Ablation study 所判斷出來重要的 feature 有哪些？
   b. 使用 Saliency map 所判斷出來重要的 feature 有哪些？
   c. 如果一起比較 categorical feature, continuous feature，你認為是公平的嗎？

Ans:

a.

Categorical Features Selected from Ablation Study: ['family support_good', 'firstborn', 'unexpectedPre', 'water_wall', 'edu_M_high']
Continuous Features Selected from Ablation Study: ['365_NO2', 'birth_weight', 'Tri2_RH', '365_CO', 'mother_age']



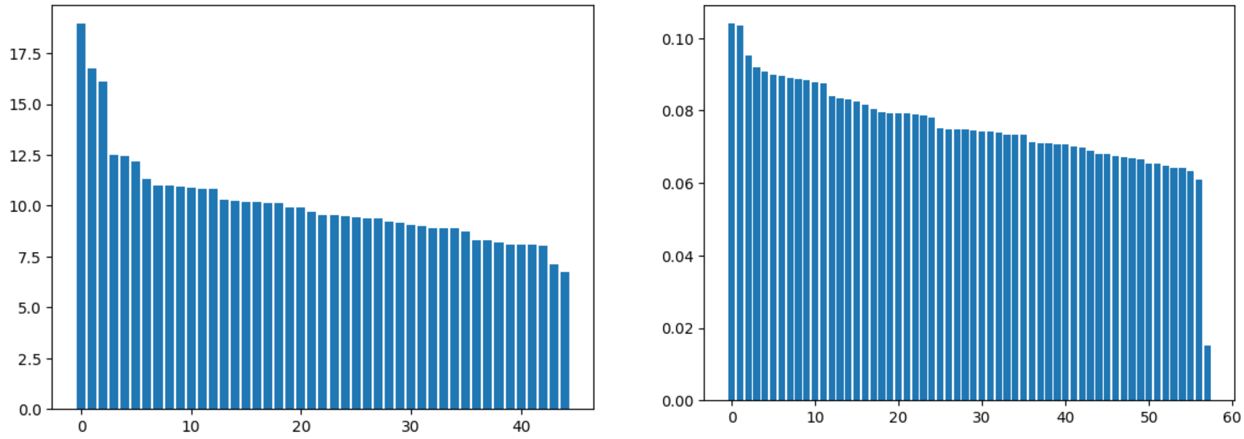


b. 如下: 因 xgboost 並無回傳 gradient 功能，只好在這邊報告單純用 simple_nn 做出來的結果。

至少在 categorical feature 上都選到相似的 feature。 Continuous feature 上差異較大。

Categorical Feature Selected from Saliency Map: ['family support_good', 'firstborn', 'mold', 'edu_M_high', 'water_wall']

Continuous Features Selected from Saliency Map: ['Tri2_O3', 'Tri1_CO', 'BW_before_D', 'APGAR_5', 'Tri3_NO']



左:category ，右: continuous

c. 不公平，以 Ablation study 為例，categorical feature 被置換造成的效果大小(AUC 改變)，和 continuous feature 改變造成的尺度本身就不一樣。例如某 continuous feature 置換成-100 ，就實際資料分布可能為為 noise 十分嚴重的輸入，對最終模型表現影響很高，但可能對 categorical feature 而言置換就成 0 不算是特別奇怪的輸入，對最終模型表現相較而言實際數值就沒相差這麼多。因此放在一起比較有失公平。