

Homework 2 Report Problem Set

Professor Pei-Yuan Wu

EE5184 – Machine Learning

學號：R07921078 系級：電機所碩一 姓名：張廷維

Problem 1. (1%) 請簡單描述你實作之 logistic regression 以及 generative model 對此 task 的表現，並試著討論可能原因。

個人測試下來的結果，當使用相同 feature 的形況下，logistic regression 通常都比 generative model 有較好的 performance，可能原因如同老師上課所說，generative model 是以某種機率分布的形況下(是有條件的)，當採用的機率分布與實際狀況不同時所得到的預測結果就會不準確。

Problem 2. (1%) 請試著討論那些 input features 中的 gender, education, martial status 等改為 one-hot encoding 進行 training process, 比較其模型準確率及其可能影響原因。

Problem 3. (1%) 請試著討論哪些 input features 的影響較大(實驗方法沒有特別限制, 但請簡單闡述實驗方法)。

由於資料後面都是每個月的資訊，因此每次 training 都會將它們加入 features，所以在我的實驗裡只討論前五項(limit_bal, sex, education, marriage, age)，實驗方式第一次先用全部 features 進行 training 後 kaggle 得到結果=0.812, 接著挑一個自己覺得最無關的 feature 掉，一開始先挑 sex，train 完之後發現比原本的 0.812 好，就在挑下一個自己覺得無關的 feature 如果 performance 變好就拿掉，如果變差就將這個 feature 保留(表是它重要)...以此類推，到最後不使用的 feature 為 sex 跟 age，做完 training 後在 kaggle 上得到結果=0.8148。

Problem 4. (1%) 請實作特徵標準化(feature normalization), 並討論其對於模型準確率的影響與可能原因。

在未使用 feature normalization 之前，不管 feature 怎麼挑都是在 0.8 以下，使用之後馬上超過 0.8，我覺得可能原因是因為同一 feature 的每一筆 data 之間的差異過大或者分布不均導致 training 出來的 model performance 不佳。

Problem 5. (1%)

PAGE

DATE

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty$$

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \stackrel{\text{Let}}{=} F$$

$$\text{Let } u = \frac{x-\mu}{\sigma} \Rightarrow du = \frac{dx}{\sigma} \Rightarrow dx = \sigma du$$

$$F = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}u^2} \sigma du$$

$$F^2 = \left(\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du \right) \left(\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}v^2} dv \right)$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(u^2+v^2)} du dv$$

$$\text{Let } 0 \leq \theta \leq 2\pi, \quad 0 \leq r < \infty$$

$$u = r \cos \theta, \quad v = r \sin \theta$$

$$F^2 = \frac{1}{2\pi} \int_0^{2\pi} \int_0^{\infty} e^{-\frac{1}{2}(r^2 \cos^2 \theta + r^2 \sin^2 \theta)} |J| dr d\theta$$

$$\text{其中 } |J| = \begin{vmatrix} \frac{\partial u}{\partial \theta} & \frac{\partial u}{\partial r} \\ \frac{\partial v}{\partial \theta} & \frac{\partial v}{\partial r} \end{vmatrix} = \begin{vmatrix} -r \sin \theta & \cos \theta \\ r \cos \theta & \sin \theta \end{vmatrix}$$

$$= |(-r \sin \theta)(\sin \theta) - (r \cos \theta)(\cos \theta)|$$

$$= -r \sin^2 \theta - r \cos^2 \theta$$

$$= |-r| |\sin^2 \theta + \cos^2 \theta| = r$$

$$F^2 = \frac{1}{2\pi} \int_0^{2\pi} \int_0^{\infty} e^{-\frac{1}{2}(r^2 (\cos^2 \theta + \sin^2 \theta))} r dr d\theta$$

$$= \frac{1}{2\pi} \int_0^{2\pi} \left[\int_0^{\infty} e^{-\frac{1}{2}r^2} r dr \right] d\theta$$

$$\text{Let } w = r^2 \Rightarrow dw = 2r dr \Rightarrow r dr = \frac{dw}{2}$$

$$F^2 = \frac{1}{2\pi} \int_0^{2\pi} \left[\int_0^{\infty} e^{-\frac{1}{2}w} \frac{dw}{2} \right] d\theta$$

$$= \frac{1}{2\pi} \int_0^{2\pi} \frac{1}{2} \left[-2e^{-\frac{1}{2}w} \right]_0^{\infty} d\theta$$

$$= \frac{1}{2\pi} \int_0^{2\pi} - \left[e^{-\frac{1}{2}w} \right]_0^{\infty} d\theta$$

$$\text{其中 } \left[e^{-\frac{1}{2}w} \right]_0^{\infty} = [0 - 1]$$

$$\text{所以 } F^2 = \frac{1}{2\pi} \int_0^{2\pi} -(-1) d\theta$$

$$= \frac{1}{2\pi} \int_0^{2\pi} d\theta = \frac{1}{2\pi} \theta \Big|_0^{2\pi}$$

$$= \frac{1}{2\pi} [2\pi - 0] = 1$$