

ML HW2 Report

b04504042 電機三 劉家豪

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

	public	private
logistic	0.76300	0.76360
generative	0.81300	0.81080

兩個都是使用所有feature，且都沒有使用normalization，從分數來看，generative好蠻多的，推測應該是因為數據量蠻大的，所以使用generative model 能達到還算不錯的效果。但如果logistic model 做一些調整，像是normalization，one-hot encoding，就會比generative好。

Problem 2. (1%) 請試著將 input feature 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。

	public	private
沒one-hot	0.81740	0.81440
有one-hot	0.81980	0.82180

有one-hot效果比較好，推測原因可能是因為one-hot是將一個feature的值做分類，所以該feature的每個分類會對應一個weight，但沒有one-hot 的，他的feature各種類只能共用同個weight。例如有個feature的值分為1~10，每個值代表一種狀態，但其實值的大小並無意義，如果共用一個weight那值的大小就會造成影響。

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)。

我用generative model 來比較 pay_0 到 pay_6這些feature的影響。

我的方法是保留pay以外的feature，然後每次都從pay_0到pay_6挑一個當feature，以下是結果：

	public	private
pay_0	0.81260	0.80880
pay_2	0.80020	0.79460
pay_3	0.78860	0.78940
pay_4	0.78760	0.78600
pay_5	0.78780	0.78560
pay_6	0.78340	0.77980

可以觀察到，對準確率來說，pay_0 > pay_2 > pay_3 > pay_4 > pay_5 > pay_6

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。

	public	private
沒normalize	0.76300	0.76360
有normalize	0.81740	0.81440

有normalize效果好很多，原因可能是因為normalize後，每個feature的scale都一樣，在做gradient descent的時候能比較快且比較容易走到好的位置。

Problem 5. (1%)

Problem 5

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

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

$$= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2\sigma^2}} d\frac{x}{\sqrt{2}\sigma}, \quad \text{let } t = \frac{x-\mu}{\sqrt{2}\sigma}$$

$$= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-t^2} dt \quad \left(\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi} \right)$$

$$= 1$$

Problem 6

Problem 6. (1%)

Problem 6

(a) $\frac{\partial E}{\partial z_k} = \frac{\partial E}{\partial y_k} \frac{\partial y_k}{\partial z_k} = \frac{\partial E}{\partial y_k} \frac{dg(z_k)}{dz_k}$

(b) $\frac{\partial E}{\partial z_j} = \frac{\partial E}{\partial z_k} \frac{\partial z_k}{\partial z_j} = \frac{\partial E}{\partial y_k} \frac{dg(z_k)}{dz_k} w_{jk} \frac{dg(z_j)}{dz_j}$

$$= w_{jk} \frac{\partial E}{\partial y_k} \frac{dg(z_k)}{dz_k} \frac{dg(z_j)}{dz_j}$$

(c) $\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial z_j} \frac{\partial z_j}{\partial w_{ij}} = w_{jk} \frac{\partial z_j}{\partial w_{ij}} \frac{\partial E}{\partial y_k} \frac{dg(z_k)}{dz_k} \frac{dg(z_j)}{dz_j}$