Tutorial 4

Machine Learning and Big Data for Economics and Finance

List of activities

I. Complete **Section 4.6 Lab: Logistic Regression, LDA, QDA, and KNN**, subsections 4.6.1, 4.6.2 and 4.6.5.

II. Complete the list of exercises in this tutorial.

Exercise 1. Logistic and logit transformations

- Show step by step that the inverse of
$$f(x)=\log\left(\frac{x}{1-x}\right)$$
is given by
$$f^{-1}(x)=\frac{1}{1+\mathrm{e}^{-x}}.$$
- Show that f^{-1} is increasing.
- Show that as $x\to-\infty$, $f^{-1}(x)\to 0$ and as $x\to\infty$, $f^{-1}(x)\to 1$.

Solution.

- $$\begin{aligned} u &=_{(i)} \log\left(\frac{x}{1-x}\right)\\ -u &=_{(\mathrm{ii})} \log\left(\frac{1-x}{x}\right)\\ \mathrm{e}^{-u} &= \frac{1-x}{x}\end{aligned}$$
where (i) follows from $\log\left(\frac{a}{b}\right)=\log(a)-\log(b)=-(-\log(a)+\log(b))=-\log\left(\frac{b}{a}\right)$.
(ii) follows from taking exponentials on both sides.
Finally, some algebra shows that $x\mathrm{e}^{-u}=1-x\Rightarrow x(1+\mathrm{e}^{-u})=1$ which finally implies that
$$x=\frac{1}{1+\mathrm{e}^{-u}}$$
which is the inverse function
- If x increases then $-x\searrow$ then $\mathrm{e}^{-x}\searrow$ then $1+\mathrm{e}^{-x}\searrow$ then $\frac{1}{1+\mathrm{e}^{-x}}$ increases.
- $x\to-\infty$ then $\mathrm{e}^{-x}\to\infty$ that is $f^{-1}(x)\to 0$.
$x\to\infty$ then $\mathrm{e}^{-x}\to 0$ then $f^{-1}(x)\to 1$.

Exercise 2. Write an `R` function `loglik_logit` that takes data and a parameter β as input and that outputs the logarithm of the likelihood of the logistic regression model.

Test your function on the dataset in the file `LR1.csv` where the model is
$$\operatorname{logit}(\Pr\{Y=1|X=x\})=-5+x\beta$$
Maximize the likelihood and compare to the function `glm`.

Solution.

Let $p(x;\beta)=\frac{\mathrm{e}^{-5+x\beta}}{1+\mathrm{e}^{-5+x\beta}}$, then we need to write the logarithm of the function
$$L(\beta)=\prod_{i:y_i=1}p(x_i;\beta)\prod_{i':y_{i'}=0}(1-p(x_{i'};\beta))$$
Taking logarithms
$$\operatorname{Log}(L(\beta))=\sum_{i:y_i=1}\log(p(x_i;\beta))+\sum_{i':y_{i'}=0}\log(1-p(x_{i'};\beta))$$
This is the function we need to code in `R`. Look at the code file attached to the solutions.