BA Individual Assignment

BA-1-1 Wenbin Zhou （周文斌）

August 14, 2021

1 Concept checks

Question a

Solution: Economically speaking, given $\theta_i = 1$, when x_i increases by 1 unit, then the log odds ratio, which is located on the very left side of the equation

$$\log \frac{\mathbb{P}(y_i = 1|\boldsymbol{x})}{\mathbb{P}(y_i = 0|\boldsymbol{x})} = \boldsymbol{\theta}^T \boldsymbol{x} + \theta_0 = \sum_{i=1}^{N} \theta_i x_i + \theta_0$$

increases by 1 unit.

Question b

Solution: It is determined by the value of k, which is the number of closest points to be identified in the model. Choosing a smaller k yields higher degree of freedom.

Question c

Solution: LASSO tends to eliminate parameters close to zero, while ridge regression tends to reduce the values but not to zero. So when fitting a linear model whose parameters are **sparse**, it is better to use LASSO, which is generally called "feature selection". In this way, we set the estimation of the parameters that are close to zero to be zero, keeping only the parameters that are worth noticing, saving time in calculation. But when the parameters are generally large in value, it is better to use ridge regression, because it converges faster than LASSO (intuitively, the slope of a quadratic function is steeper than the absolute value function at a point far from the origin).

2 Predicting mortgage defaults

This part is going to be written in the .ipynb file in the folder that contains this PDF. Please go check it out!