ING2 EILCO – Introduction to Machine Learning

Practice Exam

Augustin Cosse

November 2022

Question 1 (Fall 2022)**.** *We consider a simple regression model with two coefficients* $t = \beta_1 \tilde{x}_1^s + \beta_2 \tilde{x}_2^s$. *We assume that the data has been centered so that the model is learned on* $\tilde{x}^{(i)} = x^{(i)} - \frac{1}{N}\sum_i x^{(i)} = x^{(i)} - \overline{x}$ *and* $\tilde{t}^{(i)} = t^{(i)} - \frac{1}{N}\sum_i t^{(i)} = t^{(i)} - \overline{t}$. *moreover, after the centering step, the* $\tilde{x}^{(i)}$ *are scaled as*

$$\tilde{x}_k^{s,(i)} = \tilde{x}_k^{(i)} \leftarrow \tilde{x}_k^{(i)}/\sigma_k$$

where σ_k^2 *is the variance associated to the* k^{th} *feature of* $\boldsymbol{x}^{(i)}$,

$$\sigma_k^2 = \frac{1}{N}\sum_{i=1}^N (x_k^{(i)} - \overline{x}_k)^2, \quad \overline{x}_k = \frac{1}{N}\sum_i x_k^{(i)}$$

1. [2pts] *Show that the normal equations in this case can read as*

$$\left[\begin{array}{cc} 1 & r_{12} \\ r_{12} & 1 \end{array}\right]\left[\begin{array}{c} \beta_1 \\ \beta_2 \end{array}\right] = \left[\begin{array}{c} \gamma_1 \\ \gamma_2 \end{array}\right]$$

   What is the expression for r_{12} *in terms of the original* $x_1^{(i)}, x_2^{(i)}$ *? (Start by writing the expression of* r_{12} *as a function of the* $\tilde{x}^{s,(i)}$ *then replace the* $\tilde{x}^{s(i)}$ *by their expression as a function of the* $x^{(i)}$*)*

2. [2pts] *Give the expression of the inverse* $(\boldsymbol{X}^T\boldsymbol{X})^{-1}$ *as a function of* r_{12}. *What are the values of* r_{12} *for which this inverse is well defined ?*

Question 2 (Fall 2022)**.** *We consider the neural network shown in Fig. 2 which consists of alternating* 2 *units and* 1 *unit hidden layers. The weights associated to the* i^{th} *unit in layer* k *are denoted as* $w_{ij}^{(k)}$ *and each neuron is equipped with a sigmoid activation and a bias* $w_{i0}^{(k)}$ *(not represented on the Figure)*

1. [1pts] *Sketch the sigmoid activation*
2. [2pts] *Give the detailed expression of* $y(\boldsymbol{x}; W)$ *as a function of* $\boldsymbol{x}$, *and the* $w_{ij}^{(k)}$.
3. [4pts] *Using backpropagation, derive the gradient with respect to* $w_{11}^{(1)}$ *for a general* t *and* x *(give all the steps)*

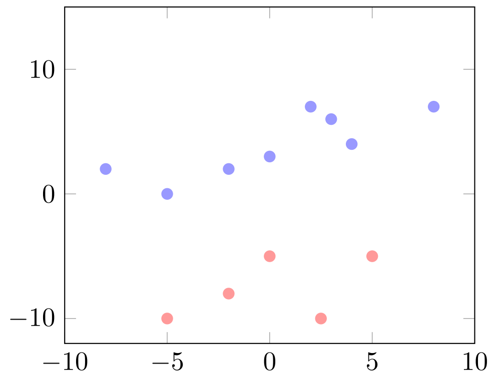

FIGURE 1 – Training set for Question 3.

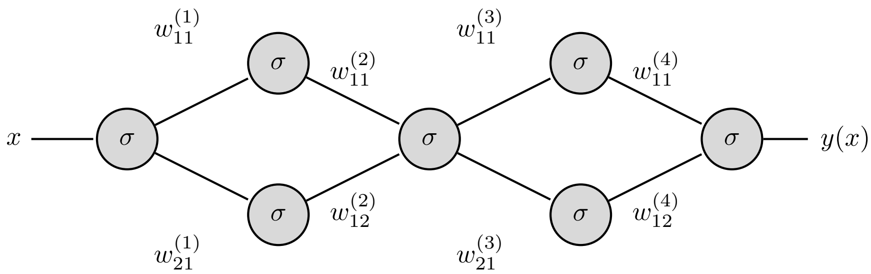

FIGURE 2 – Neural Network for Question 2

Question 3 (Fall 2022)**.** *We consider the logistic regression classifier*

$$p(t(\boldsymbol{x}) = 1|\boldsymbol{x}) = \sigma(\beta_0 + \beta_1 x_1 + \beta_2 x_2)$$
$$p(t(\boldsymbol{x}) = 0|\boldsymbol{x}) = 1 - \sigma(\beta_0 + \beta_1 x_1 + \beta_2 x_2)$$

where $\sigma(x)$ denotes the usual sigmoid function. Given the data shown in Fig. 1,

1. [2pts] *What would be a good choice for the parameters $\beta_0, \beta_1, \beta_2$ (the choice does not need to be optimal)*
2. [2pts] *Let us assume that your solution corresponds to the minimum of a certain loss $\ell(\boldsymbol{\beta})$. How would this solution change if we now decided to minimize $\ell + \lambda R(\boldsymbol{\beta})$ where R denotes the Ridge regularizer. Motivate your answer.*

Question 4 (Summer 2022)**.** *We collect data for a group of students in a machine learning class with variables*
$x_1 =$ "number of hours studied", $x_2 =$ "undergrad GPA" *and* $t =$ "receives an A". *We fit a logistic regression model to the data and produce estimated coefficients $\hat{\beta}_0 = -6$, $\hat{\beta}_1 = 0.05$, $\hat{\beta}_2 = 1$.*

1. *Estimate the probability that a student who studies for 40h and has an undergrad GPA of 3.5 gets an 'A' in the class.*
2. *How many hours would the student in part (a) need to study to have a 50% chance of getting an 'A' in the class ?*

Question 5 (Summer 2022). *We consider a $d = 2$ dimensional dataset with 2 pairs $\{\boldsymbol{x}_i, t_i\}_{i=1}^2$, i.e. $\boldsymbol{x}_i = (x_{i1}, x_{i2}) \in \mathbb{R}^2$. We assume that $x_{i1} = x_{i2}$ for $i = 1, 2$ as well as $t_1 + t_2 = 0$ and $\sum_{i=1}^2 x_{i1} = \sum_{i=1}^2 x_{i2} = 0$ so that the bias $\beta_0 = 0$. Answer the following questions*

a) *Write the ridge regression optimization problem in this setting [1pt]*

b) *Argue that in this setting, the ridge coefficient estimates satisfy $\hat{\beta}_1 = \hat{\beta}_2$ [1pt]*

c) *Write down the LASSO optimisation problem in this setting [1pt]*

d) *Argue that in this setting, the coefficients $\hat{\beta}_1$ and $\hat{\beta}_2$ are not unique - In other words, there are many possible solutions to the optimisation problem in c. Describe those solutions. [2pts]*

Question 6 (Summer 2022). *Suppose we have a dataset with five features explained below*

x_1	*GPA*
x_2	*IQ*
x_3	*Level (1 for college, 0 for High School)*
x_4	*Interaction between GPA and IQ*
x_5	*Interaction between GPA and Level*

The target is "starting salary after graduation(in thousands of dollars)". Suppose that we use a least squares approach to learn the model and got $\hat{\beta}_0 = 50$, $\hat{\beta}_1 = 20$, $\hat{\beta}_2 = .07$, $\hat{\beta}_3 = 35$, $\hat{\beta}_4 = .01$ and $\hat{\beta}_5 = -10$. Indicate whether the following are true or false

True / False	*For a fixed value of IQ and GPA, high school graduates earn more, on average, than college graduates*
True / False	*For a fixed value of IQ and GPA, college graduates earn more, on average, than high school graduates*
True / False	*For a fixed value of IQ and GPA, high school graduates earn more, on average, than college graduates provided that the GPA is high enough*
True / False	*For a fixed value of IQ and GPA, college graduates earn more, on average, than high school graduates provided that the GPA is high enough*