Assignment 5

Nayan Man Singh Pradhan

Excercise 1. (Mathematics recap)

In this problem, you will need to remind yourself about the definition of a positive definite matrix. Moreover you will look for local maxima and minima of a function of two variables.

a) Compute for the following matrices, whether they are positive definite.

$$M = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}, \quad N = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$$

b) You are given the following function:

$$f(x, y) = 3x^2y + y^3 - 3x^2 - 3y^2 + 2$$

Use the Knowledge 4.1 statement from the lecture notes to find and classify all locations of minima and maxima of this function.

(4 Points)

a) In order to be a positive definite matrix, all eigenvalues of the matrix must be positive.

For matrix $M = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$

$$\det \begin{bmatrix} 1-\lambda & 2 \\ 2 & 1-\lambda \end{bmatrix} = (1-\lambda)^2 - 4$$

$$0 = 1 - 2\lambda + \lambda^2 - 4$$

$$0 = \lambda^2 - 2\lambda - 3$$

$$\lambda_1 = 3 \qquad \lambda_2 = -1$$

$\therefore$ M is not a positive definite matrix because all its eigenvalues are not positive.

For matrix $N = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$

$$\det\begin{pmatrix} 2-\lambda & -1 & 0 \\ -1 & 2-\lambda & -1 \\ 0 & -1 & 2-\lambda \end{pmatrix} = (2-\lambda)^2(2-\lambda) + 0 + 0 - 0 - (1)(2-\lambda) - (1)(2-\lambda)$$
$$= (4-4\lambda+\lambda^2)(2-\lambda) - (2-\lambda)2$$
$$= (2-\lambda)(4-4\lambda+\lambda^2-2)$$
$$= (2-\lambda)(\lambda^2-4\lambda+2)$$
$$= (2-\lambda)\left((2+\sqrt{2})-\lambda\right)\left((2-\sqrt{2})-\lambda\right)$$

$\therefore \lambda_1 = 2$ or, $\lambda_2 = 2+\sqrt{2} \approx 3.4142$ or, $\lambda_3 = 2-\sqrt{2} \approx 0.5857$

$\therefore$ N is a positive definite matrix because all eigenvalues of N are positive.

b) $f(x,y) = 3x^2y + y^3 - 3x^2 - 3y^2 + 2$

$$\nabla f = \begin{pmatrix} 6xy - 6x \\ 3x^2 + 3y^2 - 6y \end{pmatrix}$$

$6xy - 6x = 0$ $\qquad$ $3x^2 + 3y^2 - 6y = 0$

Critical points:

	$6xy-6x$	$3x^2+3y^2-6y$
$(1,1)$	0	0
$(0,0)$	0	0
$(-1,1)$	0	0
$(0,2)$	0	0

$\therefore$ The critical points are $(0,0), (1,1), (-1,1), (0,2)$

Now, we compute the Hessian function,

$$H_{f(x)} = \begin{pmatrix} 6y-6 & 6x \\ 6x & 6y-6 \end{pmatrix} = \begin{pmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{pmatrix}$$

At $(0,0)$,

$$f_{xx} = -6, \quad f_{xy} = 0, \quad f_{yx} = 0, \quad f_{yy} = -6$$

$\Rightarrow D = f_{xx} \cdot f_{yy} - (f_{xy})^2 = (-6)(-6) - 0 = 36 > 0$

and $f_{xx} = -6 < 0$

$\therefore$ local maximum at $(0,0)$ //

At $(1,1)$

$$f_{xx} = 0, \quad f_{xy} = 6, \quad f_{yx} = 6, \quad f_{yy} = 0$$

$\Rightarrow D = 0 \cdot 0 - (6)^2 = -36 < 0$

$\therefore$ Saddle point at $(1,1)$ //

At $(-1,1)$

$$f_{xx} = 0, \quad f_{xy} = -6, \quad f_{yx} = -6, \quad f_{yy} = 0$$

$\Rightarrow D = 0 \cdot 0 - (+6)(+6) = -36 < 0$

$\therefore$ Saddle point at $(-1,1)$ //

At $(0,2)$

$$f_{xx} = 6, \quad f_{xy} = 0, \quad f_{yx} = 0, \quad f_{yy} = 6$$

$\Rightarrow D = 6 \times 6 - 0 \times 0 = 36 > 0$

and $f_{xx} = 6 > 0$

$\therefore$ Local minima at $(0,2)$ //

Excercise 2. (Linear regression by least squares)

You are given the training data $\mathcal{T} = \{((1,1)^\top, 2), ((1,2)^\top, 3), ((2,2)^\top, 3), ((2,4)^\top, 4)\}$.

a) Use linear regression by least squares to compute a predictor for the output. Do this by hand, i.e. don't use the computer and show all computation steps.

b) Predict the output for $\mathbf{x} = (1.5, 1.5)^\top$.

(4 Points)

a)

i) Construct matrix X and vector y

$$X = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 2 \\ 1 & 2 & 2 \\ 1 & 2 & 4 \end{pmatrix} \quad \text{and} \quad y = \begin{pmatrix} 2 \\ 3 \\ 3 \\ 4 \end{pmatrix}$$

ii) $b = X^T y$

$$b = \underbrace{\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 2 & 2 \\ 1 & 2 & 2 & 4 \end{pmatrix}}_{3\times4} \underbrace{\begin{pmatrix} 2 \\ 3 \\ 3 \\ 4 \end{pmatrix}}_{4\times1} = \underbrace{\begin{pmatrix} 12 \\ 19 \\ 30 \end{pmatrix}}_{3\times1}$$

iii) $A = X^T X$

$$A = \underbrace{\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 2 & 2 \\ 1 & 2 & 2 & 4 \end{pmatrix}}_{3\times4} \underbrace{\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 2 \\ 1 & 2 & 2 \\ 1 & 2 & 4 \end{pmatrix}}_{4\times3} = \underbrace{\begin{pmatrix} 4 & 6 & 9 \\ 6 & 10 & 15 \\ 9 & 15 & 25 \end{pmatrix}}_{3\times3}$$

iv) Solve

$$A \cdot \hat{\beta} = b$$

$$\begin{pmatrix} 4 & 6 & 9 \\ 6 & 10 & 15 \\ 9 & 15 & 25 \end{pmatrix} \cdot \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = \begin{pmatrix} 12 \\ 19 \\ 30 \end{pmatrix}$$

Using gaussian elimination,

$$\left(\begin{array}{ccc|c} 4 & 6 & 9 & 12 \\ 6 & 10 & 15 & 19 \\ 9 & 15 & 25 & 30 \end{array}\right) \Rightarrow \left(\begin{array}{ccc|c} 4 & 6 & 9 & 12 \\ 0 & 1 & 3/2 & 1 \\ 0 & 3/2 & 19/4 & 3 \end{array}\right) \Rightarrow \left(\begin{array}{ccc|c} 4 & 6 & 9 & 12 \\ 0 & 1 & 3/2 & 1 \\ 0 & 0 & 5/2 & 3/2 \end{array}\right)$$

$$\frac{5}{2}\beta_3 = \frac{3}{2}$$
$$\beta_3 = 0.6$$

$$\beta_2 + \frac{3}{2} \times 0.6 = 1$$
$$\beta_2 = 0.1$$

$$4\beta_1 + 6\beta_2 + 9\beta_3 = 12$$
$$4\beta_1 + 0.6 + 5.4 = 12$$
$$\beta_1 = 1.5$$

$$\therefore \hat{\beta} = \begin{pmatrix} 1.5 \\ 0.1 \\ 0.6 \end{pmatrix}$$

$$\therefore f_{\hat{\beta}}(x) = 1.5 + 0.1X_1 + 0.6X_2$$ // Ans

b) For $x = (1.5, 1.5)^T$

$$f_{\hat{\beta}}(x) = 1.5 + (0.1 \times 1.5) + (0.6 \times 1.5)$$
$$= 2.55$$ // Ans

Excercise 3. (Bayes classifier)

Prove Theorem 3.2 from the lecture. If you consider this as too hard, consult the literature from this field to solve the problem. In case you use external literature, make sure that you give a citation for that literature (i.e. say where you took it from) and rephrase the proof appropriately.

(4 Points)

Theorem 3.2 Let $X, G \in \{1, \ldots, r\} =: R_G$ be input and qualitative output with joint mixed density $\rho(x, g)$. The function $f : \mathbb{R}^D \to R_G$ that minimizes (under appropriate conditions) the expected prediction error $\mathrm{EPE}(f)$ with respect to the $0-1$ loss is given

$$f(x) = \arg\min_{g \in R_G} [1 - p(g|x)] ,$$

which can be simplified to

$$f(x) = \mathfrak{g} \quad \text{if} \quad p(\mathfrak{g}|x) = \max_{g \in R_G} p(g|x)$$

This minimizer is the **Bayes classifier**.

References:

https://machinelearning.tf.fau.de/pdfs/mlisp/Lecture4.pdf
https://www.cs.helsinki.fi/u/jkivinen/opetus/iml/2013/Bayes.pdf

Bayes classifier is a classifier that minimizes a certain probabilistic error measure. The Theorem (3.2) considers $X, G \in \{1, \ldots, r\} =: R_G$. Hence, let us assume objects to be classified from a finite set X. The set of classes is G, a classifier is then a function $f: X \to G$.

Now, let us fix an error function $E: G \times G \to \mathbb{R}$, which is the Expected Prediction Error (EPE) with respect to the 0-1 loss, such that $E(g,g) = 0$ for all g and $E(g,g') = 1$ if $g = g'$. Suppose, we take the probability distribution $P(X,G)$ over $X \times G$. Then, the Expected Prediction Error of the classifier would be:

$$\mathrm{Error}(f) = \sum_{(x,g) \in X \times G} P(x,g) \cdot E(g, f(x))$$

The Bayes Classifier minimizes the error such that:

$$f(x) = \arg\min_{f} \mathrm{Error}(f)$$

We can also write this as:

$$\text{Error}(f) = \sum_{x \in X} H_x(f(x))$$

Where,

$$H_x(g) = \sum_{g' \in G} P(x, g')\, E(g', g)$$

Then,

$$H_x(g) = \sum_{g' \neq G} P(x, g')$$

We get,

$$f(x) = \underset{g \in G}{\arg\min} \sum_{g' \neq g} P(x, g') = \underset{g \in G}{\arg\max}\, P(x, g)$$

Converting into conditional probabilities

$$f(x) = \underset{g \in G}{\arg\max}\, P(x \mid g)\, P(g)$$

Proved $\square$