Aleksei Maliutin
1224****
a*****n@student.uva.nl

Homework Assignment 1
Machine Learning 1, 18/19

2018-12-16

1 Basic Linear Algebra and Derivatives

Question 1.1

Let $\mathbf{A} = \begin{bmatrix} 2 & 3 & 4 \\ 3 & 4 & 5 \\ 4 & 5 & 6 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$, $b = [1, 2, 5]^T$

Answer the following questions with enough intermediate steps to show you did not just simply use a calculator.

a) Compute **AB**

$$\mathbf{AB} = \begin{bmatrix} 2\cdot1+3\cdot0+4\cdot1 & 2\cdot1+3\cdot1+4\cdot0 & 2\cdot0+3\cdot1+4\cdot1 \\ 3\cdot1+4\cdot0+5\cdot1 & 3\cdot1+4\cdot1+5\cdot0 & 3\cdot0+4\cdot1+5\cdot1 \\ 4\cdot1+5\cdot0+6\cdot1 & 4\cdot1+5\cdot1+6\cdot0 & 4\cdot0+5\cdot1+6\cdot1 \end{bmatrix} = \begin{bmatrix} 6 & 5 & 7 \\ 8 & 7 & 9 \\ 10 & 9 & 11 \end{bmatrix}$$

b) Are **A** and **B** invertible? How do you test for it? If so, calculate the inverses. Provide details only for your analysis of **A**.

According to Invertible Matrix Theorem, for being invertible matrix should meet some set of conditions, one of them "The determinant of **A** is not zero.". Hence, I've decided to check that condition first:

$$\det \mathbf{A} = \begin{vmatrix} 2 & 3 & 4 \\ 3 & 4 & 5 \\ 4 & 5 & 6 \end{vmatrix} = 2\cdot4\cdot6+3\cdot5\cdot4+4\cdot3\cdot5-4\cdot4\cdot4-2\cdot5\cdot5-3\cdot3\cdot6 = 48+60+60-64-50-54 = 0$$

$$\det \mathbf{B} = \begin{vmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{vmatrix} = 1\cdot1\cdot1+0\cdot0\cdot0+1\cdot1\cdot1-0\cdot1\cdot1-0\cdot1\cdot1-1\cdot0\cdot0 = 1+0+1-0-0-0 = 2$$

Calculate inverse for **B** via Gaussian elimination:

$$\left[\begin{array}{ccc|ccc} 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \end{array}\right] \to \left[\begin{array}{ccc|ccc} 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & -1 & 1 & -1 & 0 & 1 \end{array}\right] \to \left[\begin{array}{ccc|ccc} 1 & 0 & -1 & 1 & -1 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & -1 & 1 & -1 & 0 & 1 \end{array}\right] \to \left[\begin{array}{ccc|ccc} 1 & 0 & -1 & 1 & -1 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 2 & -1 & 1 & 1 \end{array}\right] \to$$

$$\to \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 0.5 & -0.5 & 0.5 \\ 0 & 1 & 0 & 0.5 & 0.5 & -0.5 \\ 0 & 0 & 1 & -0.5 & 0.5 & 0.5 \end{array}\right] \Longrightarrow \mathbf{B}^{-1} = \begin{bmatrix} 0.5 & -0.5 & 0.5 \\ 0.5 & 0.5 & -0.5 \\ -0.5 & 0.5 & 0.5 \end{bmatrix}$$

c) Is **AB** invertible? Why (not)?

Again, using the determinant of **AB** as a base condition for invertibility:

$$\det \mathbf{AB} = \begin{vmatrix} 6 & 5 & 7 \\ 8 & 7 & 9 \\ 10 & 9 & 11 \end{vmatrix} = 6\cdot7\cdot11+5\cdot9\cdot10+7\cdot8\cdot9-7\cdot7\cdot10-6\cdot9\cdot9-5\cdot8\cdot11 = 462+450+504-490-486-440 = 0$$

hence, it is not invertible.

d) Compute the solution set for the systems Bx = b and Ax = b. What can we say about the second system due to the invertibility property you determined previously?

Taking into account the simplicity of **B** and **b**, we can use Cramerś rule for finding the solution for the given system. For that purpose, we should change i-th column in matrix **B** on the column **b** and calculate determinants:

$$\Delta_1 = \det B_1 = \begin{vmatrix} 1 & 1 & 0 \\ 2 & 1 & 1 \\ 5 & 0 & 1 \end{vmatrix} = 1\cdot1\cdot1+2\cdot0\cdot0+1\cdot5\cdot1-0\cdot1\cdot5-2\cdot1\cdot1-1\cdot1\cdot0 = 1+0+5-0-2-0 = 4$$

$$\Delta_2 = \det B_2 = \begin{vmatrix} 1 & 1 & 0 \\ 0 & 2 & 1 \\ 1 & 5 & 1 \end{vmatrix} = 1\cdot2\cdot1+0\cdot0\cdot5+1\cdot1\cdot1-0\cdot2\cdot1-0\cdot1\cdot1-1\cdot1\cdot5 = 2+0+1-0-0-5 = -2$$

$$\Delta_2 = \det B_2 = \begin{vmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 1 & 0 & 5 \end{vmatrix} = 1\cdot1\cdot5+0\cdot0\cdot1+1\cdot2\cdot1-1\cdot1\cdot1-0\cdot1\cdot5-1\cdot2\cdot0 = 5+0+2-1-0-0 = 6$$

Hence, taking into account that $\Delta = \det \mathbf{B} = 2$, the solution for system is following:
$x = [\Delta_1/\Delta, \Delta_2/\Delta, \Delta_3/\Delta]^\top = [2, -1, 3]^\top$

According to Rouché–Capelli theorem, the second system has NO solutions, because the rank of **A** is not equal to rank of its augmented matrix **Ab**

Question 1.2

Find the derivative of the following functions with respect to x.

a) $(\frac{2}{x^2} + x^{-7} + x^3)' = 2\cdot(-2)\cdot x^{-3} + (-7)\cdot x^{-8} + 3\cdot x^2 = -\frac{4}{x^3} - \frac{7}{x^8} + 3x^2$

b) $(xe^{-\sqrt[5]{x}})' = x'\cdot e^{-\sqrt[5]{x}} + x\cdot(e^{-\sqrt[5]{x}})' = 1\cdot e^{-\sqrt[5]{x}} + x\cdot(e^{-\sqrt[5]{x}}\cdot -\frac{1}{5}\cdot x^{-\frac{4}{5}}) = e^{-\sqrt[5]{x}}\cdot(1-\frac{1}{5}\sqrt[5]{x})$

c) $(\frac{1}{x} + \ln(x^2))' = -\frac{1}{x^2} + 2x\cdot\frac{1}{x^2}$
d) $(\sigma(x))' = (\frac{1}{1+e^{-x}})' = \frac{(0\cdot(1+e^{-x})-1\cdot(1+e^{-x})')}{(1+e^{-x})^2} = \frac{e^{-x}}{(1+e^{-x})^2}$

e) Taking into account the specific of max function, it is obvious that the derivative also is a piece-wise function:

$$max(0,x) = \begin{cases} 0 & x<0 \\ x=0 & x=0 \\ x & x>0 \end{cases} \rightarrow (max(0,x))' = \begin{cases} 0 & x<0 \\ undefined & x=0 \\ 1 & x>0 \end{cases}$$

What is the shape of the following gradients:
g) Scalar, $[1\times1]$
h) $[1\times n]$, because of convention that defines the gradient as a row-vector.
i) $[m\times n]$
Find the gradient of the following functions. Make their shapes explicit.
j) $f(x) = 2e^{x_2 - ln(x_1^{-1}) - sin(x_3x_1^2)}$
Hence, the gradient is following: $2e^{x_2 ln(x_1^{-1}) sin(x_3x_1^2)} \times [(\frac{1}{x} - \cos(x_3x_1^2)\cdot 2x_1x_3), 1, -\cos(x_3x_1^2)\cdot x_1^2]$, shape: $[1\times3]$
k) $h(y) = (g\circ f)(y) = ysin(y)^3 + e^{ycos(y)}$
Hence, the gradient is following: $[\sin^2(y)\cos y\cdot 3y^3 + \sin^3(y)\cdot 3y^2 + e^{ycos(y)}\cdot(\cos y - y\cos y)]$, shape: $[1\times1]$
l) $h(y,z) = (g\circ f)(y,z) = (ysin(y)+z)^3 + e^{ycos(y)+z^2}$

Question 1.3

The following questions are good practice in manipulating vectors and ma- trices and they are very important for solving for posterior distributions.
a) $\nabla_\mu x^T\Sigma^{-1}\mu = x^T\Sigma^{-1}$, based on 5.118 from *mathematics for machine learning* book
b) $\nabla_\mu \mu^T\Sigma^{-1}\mu = \mu^T(\Sigma^{-1} + (\Sigma^T)^{-1})$, based on 5.120 from *mathematics for machine learning* book

c) $\nabla_{\mathbf{W}} f$, where $f = \mathbf{W}x$ hence $\rightarrow \frac{\partial \mathbf{f}}{\partial \mathbf{A}} = \begin{bmatrix} \frac{\partial f_1}{\partial \mathbf{A}} \\ \frac{\partial f_2}{\partial \mathbf{A}} \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} \mathbf{x}^\top \\ \mathbf{o}^\top \end{bmatrix} \\ \begin{bmatrix} \mathbf{o}^\top \\ \mathbf{x}^\top \end{bmatrix} \end{bmatrix} \in \mathbb{R}^{1\times(2\times3)}$

2 Probability Theory

For these questions you will practice manipulating probabilities and probability density functions using the sum and product rules.

Question 2.1

Suppose you meet Bart. Bart is a quiet, introvert man in his midfourties. He is very shy and withdrawn, invariably helpful but with little interest in people or in the world of reality. A meek and tidy soul, he has a need for order and structure, and a passion for detail.
a) What do you think: Is Bart a farmer or is he a librarian?

Librarian, because having a passion for details is a bad for farmer due to the stochastic nature of the majority events in nature. Also, based on personal experience, a lot of librarians really like to know a little bit more about peoples rather than only *favorite books*.

b) Suppose we formulate our belief such that we think that Bart is a librarian, given that he is an introvert, with 0.8 probability. If he were an extrovert, that probability is 0.1. Since the Netherlands is a social country, you assume that the probability of meeting an extrovert is 0.7 Define the random variables and values they can take on, both with symbols and numerically.

Define random variables as following:
$X = \{librarian, farmer\}, Y = \{extrovert, introvert\}$,
and the probabilities are
$p(X =' librarian' \mid Y =' introvert') = 0.8$
$p(X =' librarian' \mid Y =' extrovert') = 0.1$
$p(Y =' extrovert') = 0.7$

c) What is the probability that Bart is a librarian?
Calculate the *raw* probability of being librarian:
$p(X =' librarian') = p(X =' librarian' \mid Y =' introvert') \cdot p(Y =' introvert') + p(X =' librarian' \mid Y =' extrovert') \cdot p(Y =' extrovert') = 0.8 \cdot (1 - 0.7) + 0.1 \cdot 0.7 = \mathbf{0.31}$

d) Suppose you looked up the actual statistics and find out that there are 1000 times more farmers than librarians in the Netherlands. Additionally, Bart's friend is telling you that if Bart were a librarian, the probability of him being a extrovert would be low, and equal to 0.1. Given this information, how should we update our belief in that Bart is a librarian, if he is an introvert? How does does influence the answer to the previous question: What is the probability that Bart is a librarian?
According to Bayes rule we can recalculate probability
$p(X =' librarian' \mid Y =' introvert') = \frac{p(Y =' introvert' \mid X =' librarian') \cdot p(X =' librarian')}{p(Y =' introvert')} = \frac{0.9 \cdot 0.001}{0.3} = 0.003$
$p(X =' librarian' \mid Y =' extrovert') = \frac{p(Y =' extrovert' \mid X =' librarian') \cdot p(X =' librarian')}{p(Y =' extrovert')} = \frac{0.1 \cdot 0.001}{0.7} = 0.000142$
hence the new probability that Bart is a librarian $p(X =' librarian') = 0.003142$