Assignment 3

MASD 2018

Department of Computer Science
University of Copenhagen

Casper Lisager Frandsen <fsn483@alumni.ku.dk>

Version 1
Due: 23. September 2018, 10:00

Contents

Exercise 1

(a)

To find all the extremum points of the function we have to look at where $f'(x) = 0$. To find whether the points are local maxima or minima, we have to look at $f''(x)$.

$$f(x) = 2x^2 + (x-4)^3$$
$$f'(x) = 4x + 3(x-4)^2$$

First we try to find if, and where $f'(x)$ intersects the x-axis.

$$4x + 3(x-4)^2 = 0$$
$$d = \frac{-4 \pm \sqrt{16-12}}{6} = \frac{-4 \pm 2}{6}$$
$$d_1 = \frac{-1}{3}, d_2 = -1$$

Since both d_1 and d_2 are negative, there exist no intersections between $f(x)$ and the x-axis. This means that we cannot draw a conclusion whether it is a maximum or minimum.

(b)

To find all the extremum points of the function we have to look at where $f'(x) = 0$. To find whether the points are local maxima or minima, we have to look at $f''(x)$.

$$f(x) = x^2 ln(x)$$
$$f'(x) = 2x ln(x) + x$$

First we try to find if, and where $f'(x)$ intersects the x-axis. For this we used Wolfram Alpha to determine that it is $\frac{1}{\sqrt{e}}$

Now we know that $f(x)$ has a local extremum in $f(\frac{1}{\sqrt{e}}$. To find whether it is a local maximum or minimum, we simply compute the following:

$$f''(x) = 2ln(x) + 3$$
$$f''(\frac{1}{\sqrt{e}}) = 2$$

Now we know that $f''(x)$ is positive in the local extremum, and can conclude that it is a local minimum.

(c)

For this exercise we have to compute the derivatives both with respect to x_1 and x_2.

$$f(x_1, x_2) = x_1^2 + 2x_1x_2 + 3x_2^2$$

$$\frac{\partial y}{\partial x_1} = 2x_1 + 2x_2$$

$$\frac{\partial y}{\partial x_2} = 2x_1 + 6x_2$$

Then we solve for x_1 and x_2, using matrices.

$$\left[\begin{array}{cc|c} 2 & 2 & 0 \\ 2 & 6 & 0 \end{array}\right] = \left[\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \end{array}\right]$$

Thus, we can see that $f(x)$ intersects the x-axis when $x_1 = x_2 = 0$ Now, to find whether it is a local maximum or minimum, we use the eigenvalues of the Hessian matrix.

$$\frac{\partial^2 y}{\partial x_1 \partial x_2} = 2$$

$$\frac{\partial^2 y}{\partial^2 x_1} = 2$$

$$\frac{\partial^2 y}{\partial x_2 \partial x_1} = 2$$

$$\frac{\partial^2 y}{\partial^2 x_2} = 6$$

Thus, we can see that the Hessian matrix looks as follows:

$$\begin{bmatrix} 2 & 2 \\ 2 & 6 \end{bmatrix}$$

$$\begin{bmatrix} 2 & 2 \\ 2 & 6 \end{bmatrix} - \lambda I = \begin{bmatrix} 2-\lambda & 2 \\ 2 & 6-\lambda \end{bmatrix}$$

Then we find the determinants of the matrix:

$$D\begin{bmatrix} 2-\lambda & 2 \\ 2 & 6-\lambda \end{bmatrix} = \lambda^2 - 8\lambda + 8$$

Solving this with Wolfram Alpha, $\lambda^2 - 8\lambda + 8 = 0$ gives us the following:

$$\lambda_1 = 4 - 2\sqrt{2}, \lambda_2 = 2(2 + 2\sqrt{2})$$

Since both of these values are positive, we can determine that the point is a local minimum.

(d)

For this exercise we have to compute the derivatives both with respect to x_1 and x_2.

$$f(x_1, x_2) = (x_1 - x_2)^2$$
$$\frac{\partial y}{\partial x_1} = 2x_1 - 2x_2$$
$$\frac{\partial y}{\partial x_2} = -2x_1 + 2x_2$$

Then we solve for x_1 and x_2, using matrices.

$$\left[\begin{array}{rr|r} 2 & -2 & 0 \\ -2 & 2 & 0 \end{array}\right] = \left[\begin{array}{rr|r} 1 & -1 & 0 \\ 0 & 0 & 0 \end{array}\right]$$

With this we can see that $x_1 - x_2 = 0$, in other words, $f(x_1, x_2) = 0$ when $x_1 = x_2$. This means that $f(x_1, x_2)$ intersects the x-axis at all points where $x_1 = x_2$.
Now, to find whether it is a local maximum or minimum, we use the eigenvalues of the Hessian matrix.

$$\frac{\partial^2 y}{\partial x_1 \partial x_2} = -2$$
$$\frac{\partial^2 y}{\partial^2 x_1} = 2$$
$$\frac{\partial^2 y}{\partial x_2 \partial x_1} = -2$$
$$\frac{\partial^2 y}{\partial^2 x_2} = 2$$

Thus, we can see that the Hessian matrix looks as follows:

$$\begin{bmatrix} 2 & -2 \\ -2 & 2 \end{bmatrix}$$

$$\begin{bmatrix} 2 & -2 \\ -2 & 2 \end{bmatrix} - \lambda I = \begin{bmatrix} 2-\lambda & -2 \\ -2 & 2-\lambda \end{bmatrix}$$

Then we find the determinants of the matrix:

$$D \begin{bmatrix} 2-\lambda & -2 \\ -2 & 2-\lambda \end{bmatrix} = \lambda^2 - 4\lambda$$

Solving $\lambda^2 - 4\lambda = 0$ with Wolfram Alpha shows us that $\lambda_1 = 0, \lambda_2 = 4$. Since both values aren't nonzero, the test is inconclusive, and we can't tell whether it is a local maximum or minimum.

Exercise 2

(a)

Mathematically, the model calculates the length of a matrix norm, usually filled with very small numbers. The matrices A and B are defined early on as matrices that, when multiplied together, form a matrix very similar to M, but with more values where M has zeroes. These values in which M has a zero value, but AB has a value are basically predictions for what they would be in M, if the dataset was complete. When multiplying by I however, these values are negated, since we don't need them to check accuracy. Since AB and M are so close to each other, we are going to get very small values in the resulting matrix when we say $M - AB$. When we calculate the matrix norm of the resulting matrix, we are essentially adding all of the differences together, to check whether they are inside a predetermined threshold. Adding them all together and squaring them allows the algorithm to easily check for an average error rate. But if even one value in AB is significantly off, the error value that the function returns will become large easily.

The two matrices A and B can be seen as how much each person likes a genre, and how much a film scores in a specific genre respectively. In this case, there are only two general genres, which is why A and B are of the form (6×2) and (2×10). In this case with the Netflix problem, multiplying a persons like of a particular genre with how much the film scores in that particular genre, continuing to do this for all genres and adding it all together gives a good guess at how much a person would like a movie they haven't seen before.

(b)

Here, we have to show that the following is true:

$$||I \odot (M - AB)||^2 = \sum_{i=1}^{6} \sum_{j=1}^{10} I_{ij}(M_{ij} - (a_{i1}b_{1j} + a_{i2}b_{2j}))^2$$

First we start off by showing that each element in the matrix can be written as follows:

$$AB_{ij} = (a_{i1}b_{1j} + a_{i2}b_{2j})$$

First, we have to remember the rules for matrix multiplication, and recall that any element in the resulting square matrix with length m can be described as thus, when the matrices being multiplied have the forms $m \times n$ and $n \times m$.

$$a_{i1}b_{1j} + a_{i2}b_{2j} + ... + a_{in}b_{nj}$$

This can also be applied to AB, in which case we find that any element in the resulting matrix can be described as $(a_{i1}b_{1j} + a_{i2}b_{2j})$, which is exactly what we want to show.

If we recall the rules for matrix subtraction, we can also figure out that each element in the new matrix will be defined as thus, where X and Y are the matrices being subtracted from one another, with the resulting matrix Z:

$$Z_{ij} = X_{ij} - Y_{ij}$$

This must also be true for $M - AB$:

$$M_{ij} - (a_{i1}b_{1j} + a_{i2}b_{2j})$$

Since our previous operations have all been element-wise, multiplying the matrix I on element-wise is simple:

$$I_{ij}(M_{ij} - (a_{i1}b_{1j} + a_{i2}b_{2j}))$$

Now, the matrix norm is defined as being the same as one long vector V with all the same elements as the matrix:

$$||V||^2 = \sum_{i=1}^{n}(V_i^2)$$

Since a matrix can be described as a vector of vectors, the following must be true for the matrix norm, of a matrix N, with the form $n \times m$:

$$||N||^2 = \sum_{i=1}^{n}\sum_{j=1}^{m} N_{ij}$$

With this knowledge we can conclude that the matrix norm squared of $I_{ij}(M_{ij} - (a_{i1}b_{1j} + a_{i2}b_{2j}))$ must be equal to:

$$\sum_{i=1}^{6}\sum_{j=1}^{10} I_{ij}(M_{ij} - (a_{i1}b_{1j} + a_{i2}b_{2j}))^2$$

Thus, we have shown that:

$$||I \odot (M - AB)||^2 = \sum_{i=1}^{6}\sum_{j=1}^{10} I_{ij}(M_{ij} - (a_{i1}b_{1j} + a_{i2}b_{2j}))^2$$

Exercise 3

(a)

To derive the function given with respect to a_{km} first, we have to make a few important realisations:

$$\sum_{i=1}^{6}\sum_{j=1}^{10} I_{kj}(M_{kj} - (a_{k1}b_{1j} + a_{k2}b_{2j}))^2$$

Since we are deriving with a_{km}, we can describe it as relating to one row, where $i = k$. This means that if $i \neq k$ the following is true:

$$\frac{\partial E}{\partial a_{km}} I_{kj}(M_{kj} - (a_{k1}b_{1j} + a_{k2}b_{2j}))^2 = 0$$

Now we can rewrite the function to the following:

$$\frac{\partial E}{\partial a_{km}} = \sum_{j=1}^{10} I_{kj}(M_{kj} - (a_{k1}b_{1j} + a_{k2}b_{2j}))^2$$

Using the product rule, we can simply ignore the sum symbol when deriving. For this, we will derive with respect to $m \in \{1,2\}$, instead of one at a time. We will use the chain rule in this step.

$$\frac{\partial E}{\partial a_{km}} = \sum_{j=1}^{10} I_{kj} 2(M_{kj} - (a_{k1}b_{1j} + a_{k2}b_{2j})) * -(b_{1j} + b_{2j})$$

Now, since we derived with $m \in \{1,2\}$, we can see that if we had derived it with either 1 or 2, $-(b_{1j} + b_{2j})$ would have become the following:

$$-(b_{mj})$$

We can insert this into the expression instead:

$$\frac{\partial E}{\partial a_{km}} = \sum_{j=1}^{10} I_{kj} 2(M_{kj} - (a_{k1}b_{1j} + a_{k2}b_{2j})) * -(b_{mj})$$

The expression has been derived now, and we can reduce it to the following:

$$2\sum_{j=1}^{10} I_{kj}(-M_{kj}b_{mj} + a_{k1}b_{1j}b_{mj} + a_{k2}b_{2j}b_{mj})$$

This is what we wanted to show.

To derive the function given with respect to b_{ml}, we have to make the same realisation as in the first part: Since we are deriving with b_{ml}, we can describe it as relating to one row, where $j = m$. This means that if $j \neq m$ the following is true:

$$\frac{\partial E}{\partial b_{ml}} I_{il}(M_{il} - (a_{i1}b_{1l} + a_{i2}b_{2l}))^2 = 0$$

We can use this to rewrite the expression as follows:

$$\frac{\partial E}{\partial b_{ml}} = \sum_{i=1}^{6} I_{il}(M_{il} - (a_{i1}b_{1l} + a_{i2}b_{2l}))^2$$

Now we derive it, and can again use the product rule to derive each element of the sum, and the chain rule.

$$\frac{\partial E}{\partial b_{ml}} = 2\sum_{i=1}^{6} I_{il}(M_{il} - (a_{i1}b_{1l} + a_{i2}b_{2l})) * -(a_{i1} + a_{i2})$$

We can make the same realisation as before, that since we derived with $m \in \{1,2\}$, we can see that if we had derived it with either 1 or 2, $-(a_{i1} + a_{i2})$ would have become the following:

$$-(a_{im})$$

We can again put this in the function instead of $-(a_{i1} + a_{i2})$:

$$\frac{\partial E}{\partial b_{ml}} = 2\sum_{i=1}^{6} I_{il}(M_{il} - (a_{i1}b_{1l} + a_{i2}b_{2l})) * -(a_{im})$$

The expression has now been derived, and we will reduce it to the following:

$$\frac{\partial E}{\partial b_{ml}} = 2\sum_{i=1}^{6} I_{il}(-M_{il}a_{im} + a_{i1}b_{1l}a_{im} + a_{i2}b_{2l}a_{im})$$

This is also what we wanted to show.