Problem set 1: Math for Machine Learning

Jiarong Ye

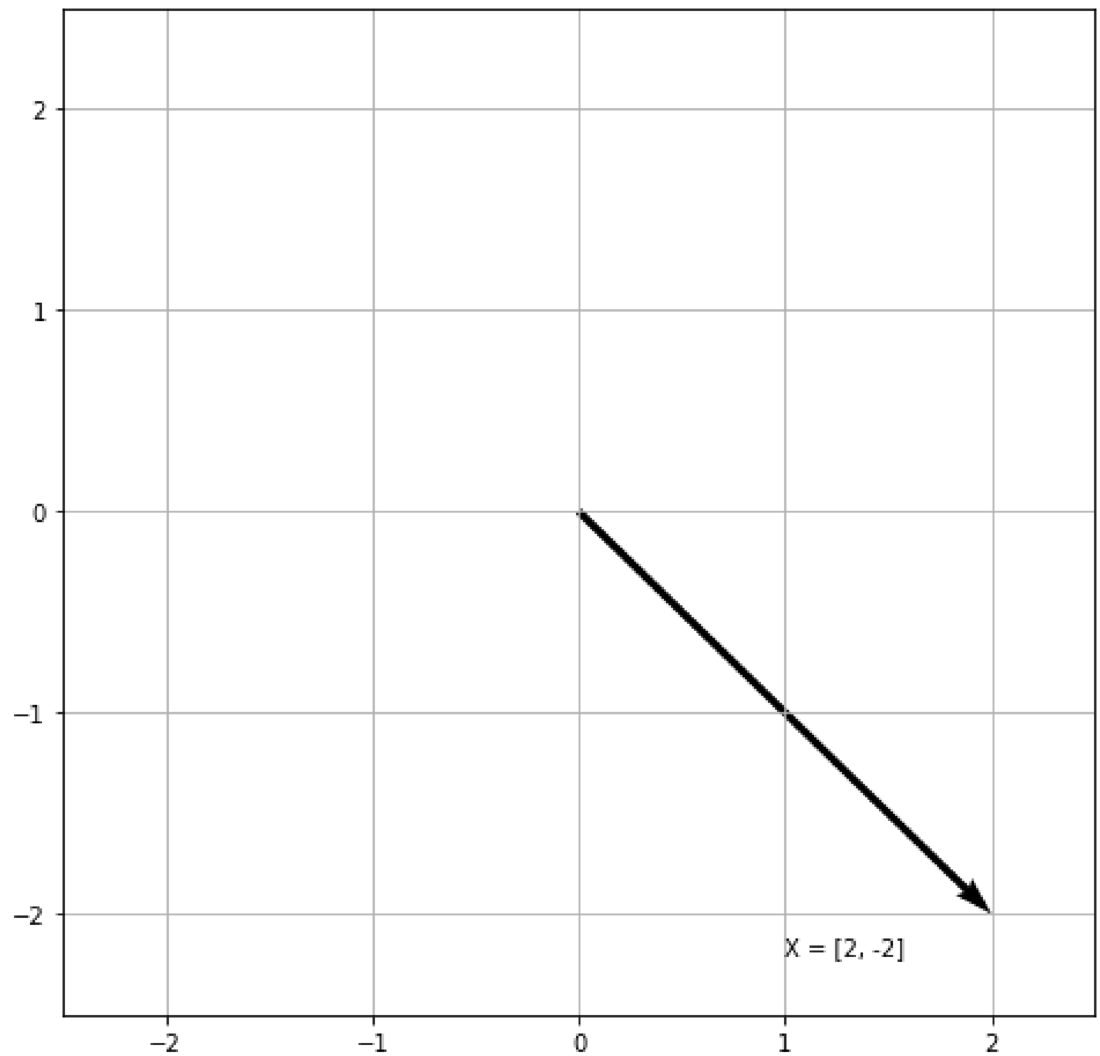
August 26, 2018

Q1

1.(20 pts.) Let X = [2, -2] be a vector.

(a) Plot the vector in a 2-dimensional Euclidean space

```
In [1]: import numpy as np
        import matplotlib.pyplot as plt
        %matplotlib inline

        def draw_vector(x):
            plt.figure(figsize=(8, 8))
            for i in x:
                X, Y, U, V = zip(*np.array(x))
                ax = plt.gca()
                ax.quiver(X, Y, U, V, angles='xy', scale_units='xy', scale=1)
                ax.annotate('X = [{}, {}] '.format(i[2], i[3]), xy=(i[2]-1, i[3]-0.2))
            ax.set_xlim([-2.5, 2.5])
            ax.set_ylim([-2.5, 2.5])
            plt.grid(True)
            plt.draw()
            plt.show()

In [2]: draw_vector([[0,0,2,-2]])
```



(b) Compute the Euclidian norm of the vector X.

The general case of Euclidian norm of a vector:

$$\|\boldsymbol{x}\|_2 = \sqrt{x_1^2 + \cdots + x_n^2}$$

Thus, the Euclidian norm of the vector X is:

$$||x||_2 = \sqrt{2^2 + (-2)^2} = 2\sqrt{2}$$

(c) Normalize the vector X to obtain a vector of unit length

$$\text{normalized X } = \frac{X}{||X||_2} = [\frac{2}{||x||_2}, \frac{-2}{||x||_2}] = [\frac{2}{2\sqrt{2}}, \frac{-2}{2\sqrt{2}}] = [\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}]$$

check the answer:

$$\text{length of normalized X } = \sqrt{\left(\frac{\sqrt{2}}{2}\right)^2 + \left(-\frac{\sqrt{2}}{2}\right)^2} = \sqrt{\frac{1}{2}+\frac{1}{2}} = 1$$

so the normalized vector X is unit length, the answer is checked.

(d) Compute the vector that is normal to X.

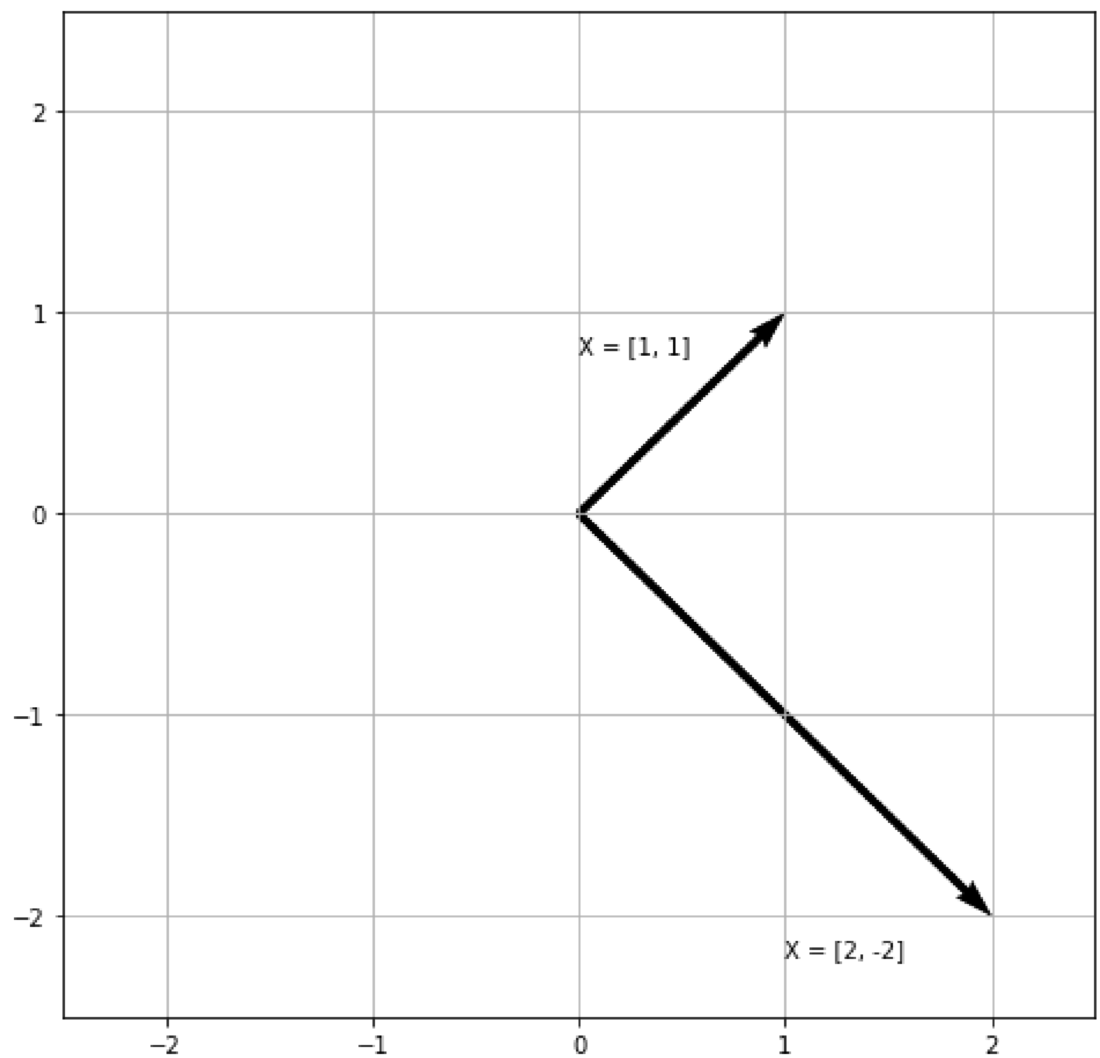
Since the vector $X = [2, -2]$ in a 2-dimensional Euclidean space could be written as a line:

$$x + y = 0$$

the normal vector of a 2-dimensional line will have the direction vector of an orthogonal line to it, and the perpendicular vector is (1,1) in this case.

So the vector that is normal to X is $[1, 1]$

```
In [3]: # check the graph:
        draw_vector([[0,0,2,-2], [0,0,1,1]])
```



(e) Compute the transpose of the vector X.

The transpose vector of X is:

$$X^T = \begin{bmatrix} 2 \\ -2 \end{bmatrix}$$

Q2

2.(20 pts.) Joe takes a lab test and the result comes back positive. The test returns a correct positive result in only 98% of the cases in which the disease is actually present, and a correct negative result in only 97% of the cases in which the disease is not present. Furthermore, .008 of the entire population have this cancer. Does Joe have cancer? Justify your answer.

$$P(cancer) = 0.008$$
$$P(no\ cancer) = 0.992$$

$$P(positive \mid cancer) = 0.98$$
$$P(negative \mid cancer) = 0.02$$

$$P(negative \mid no\ cancer) = 0.97$$
$$P(positive \mid no\ cancer) = 0.03$$

Since the result of Joe's test comes back as positive:
Thus

$$\begin{aligned} P(cancer \mid positive) &= \frac{P(positive, cancer)}{P(positive)} && (1) \\ &= \frac{P(positive \mid cancer)P(cancer)}{P(positive \mid cancer)P(cancer) + P(positive \mid no\ cancer)P(no\ cancer)} && (2) \end{aligned}$$

And since:

$$P(positive \mid cancer)P(cancer) = (0.98)(0.008) = 0.00784$$

$$\begin{aligned} &P(positive \mid cancer)P(cancer) + P(positive \mid no\ cancer)P(no\ cancer) && (3) \\ &= (0.98)(0.008) + (0.03)(0.992) = 0.0376 && (4) \end{aligned}$$

$$
\begin{aligned}
P(cancer \mid positive) &= \frac{P(positive, cancer)}{P(positive)} && (5)\\
&= \frac{P(positive \mid cancer)P(cancer)}{P(positive \mid cancer)P(cancer) + P(positive \mid no\ cancer)P(no\ cancer)} && (6)\\
&= \frac{0.00784}{0.0376} && (7)\\
&= 0.2085 && (8)
\end{aligned}
$$

With Joe's test result being positive, there's approximately 20.85% chance that Joe have cancer, so he most likely does not have cancer.

Q3

3.(20 pts.) Consider the function $f(x, y) = x^2 + xy + y^2$.

(a) Compute the first order partial derivatives $\partial f x$ and $\partial f y$

$$\partial f x = \frac{\partial(x^2 + xy + y^2)}{\partial x} = 2x + y$$

$$\partial f y = \frac{\partial(x^2 + xy + y^2)}{\partial y} = 2y + x$$

(b) Compute the Hessian (matrix of second order partial derivatives) of f.

$$H = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix} \tag{9}$$

$$\frac{\partial^2 f}{\partial x^2} = 2$$

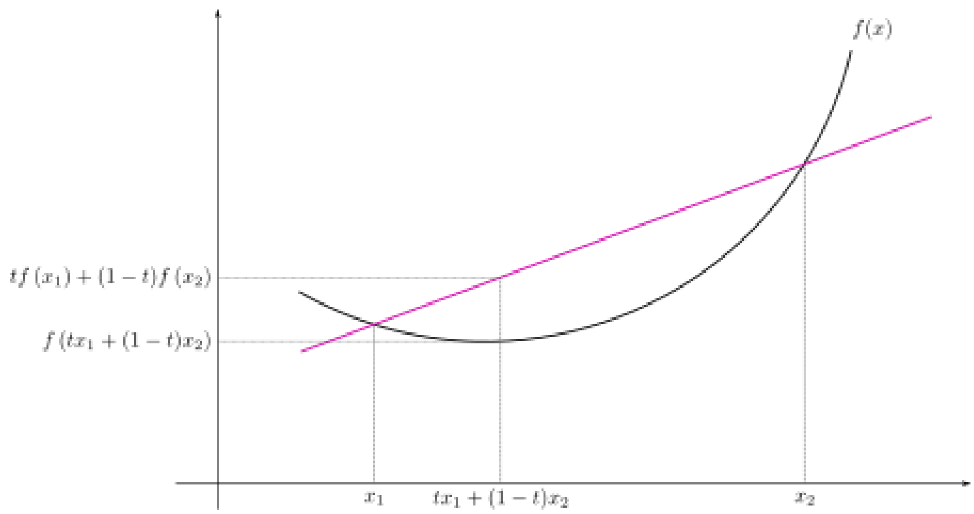
$$\frac{\partial^2 f}{\partial x \partial y} = 1$$

$$\frac{\partial^2 f}{\partial y \partial x} = 1$$

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

$$H = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \tag{10}$$

(c) Is f(x, y) a convex function of x and y? Justify your answer.



As shown in the graph, a function $f(x)$ is considered as convex on an interval $[a, b]$ if for any two points x_1 and x_2 in $[a, b]$ and any λ where $0 < \lambda < 1$

$$f[\lambda x_1 + (1-\lambda)x_2] \leq \lambda f(x_1) + (1-\lambda) f(x_2)$$

And a convenient test for convexity is to calculate its second derivative since a twice differentiable function of one variable is convex on an interval if and only if its second derivative is non-negative.

- for x , $f(x) = x^2 + xy + y^2$,
$$\frac{\partial^2 f}{\partial x^2} = 2 \geq 0$$
so $f(x, y)$ is a convex function of x
- for y , $f(y) = y^2 + yx + x^2$,
$$\frac{\partial^2 f}{\partial y^2} = 2 \geq 0$$
so $f(x, y)$ is a convex function of y

In conclusion, $f(x, y)$ a convex function of x and y

(d) Compute the values of x and y at which $f(x, y)$ is minimized.

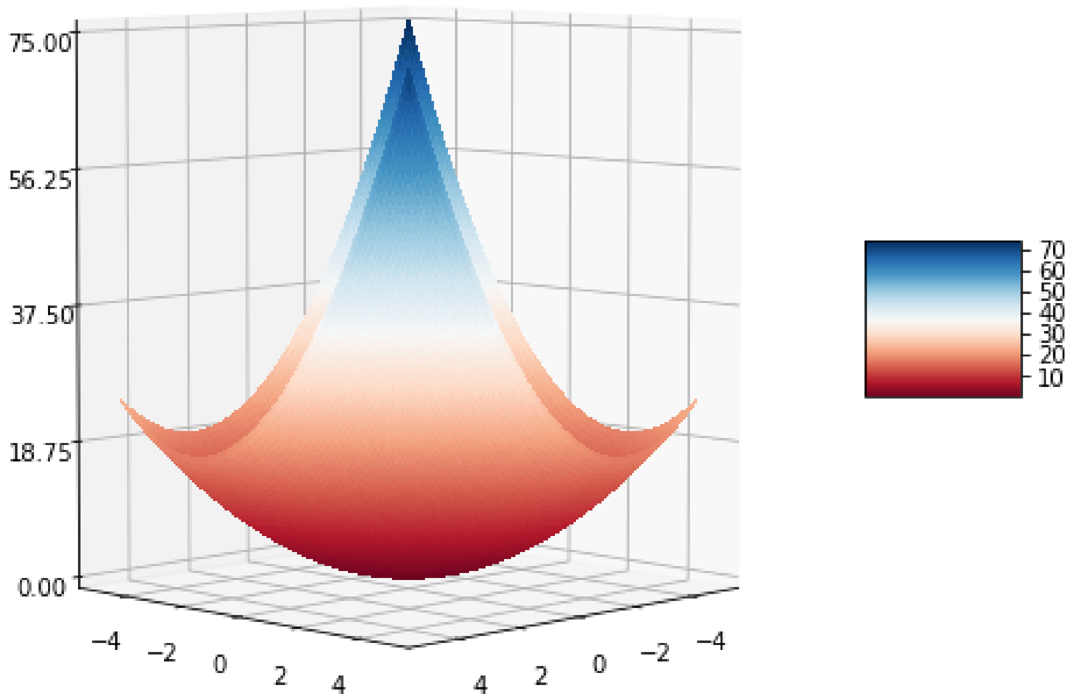
Since we already know from part (c) that $f(x, y) = x^2 + xy + y^2$ is convex for both x and y, so when $f(x, y)$ is minimized, there's only one global minimum value exists (no local extremes).
Thus, the requirement for min $f(x, y)$ is:

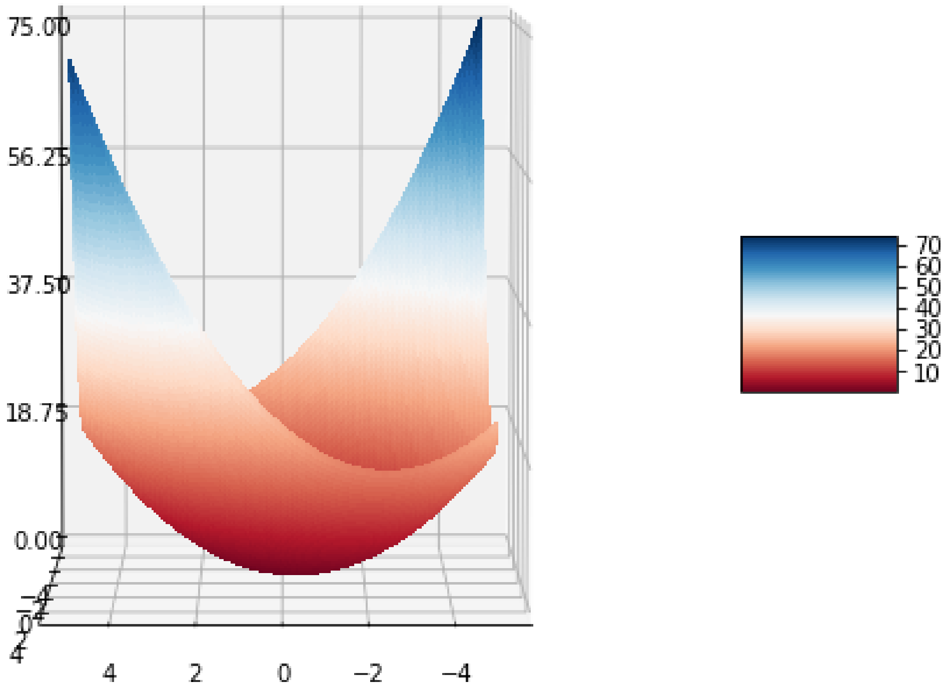
$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} = \begin{bmatrix} 2x + y \\ 2y + x \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{11}$$

So:

$$\begin{cases} x = 0 \\ y = 0 \end{cases}$$

```
In [4]: from numpy import exp,arange
        from pylab import meshgrid,cm,imshow,contour,clabel,colorbar,axis,title,show
        from mpl_toolkits.mplot3d import Axes3D
        from matplotlib import cm
        from matplotlib.ticker import LinearLocator, FormatStrFormatter
        import matplotlib.pyplot as plt

        def plot_func(angle_up, angle_side):
            x = arange(-5.0,5.0,0.1)
            y = arange(-5.0,5.0,0.1)
            X,Y = meshgrid(x, y)
            Z = X**2+X*Y+Y**2

            fig = plt.figure(figsize=(8,8))
            ax = fig.gca(projection='3d')
            surf = ax.plot_surface(X, Y, Z, rstride=1, cstride=1,
                                   cmap=cm.RdBu,linewidth=0, antialiased=False)

            ax.zaxis.set_major_locator(LinearLocator(5))
            ax.zaxis.set_major_formatter(FormatStrFormatter('%.02f'))
            ax.view_init(angle_up, angle_side)
            fig.colorbar(surf, shrink=0.5, aspect=1)
            plt.show()

In [5]: plot_func(5, 45), plot_func(5, 90)
```


75.00
56.25
37.50
18.75
0.00
−4
−2
0
2
4
4
2
0
−2
−4
70
60
50
40
30
20
10



Q4

4.(20 pts.) Suppose a coin is tossed. The observed outcomes are H,T,H,H,H,T,H,H,T,H,H,T (where H denotes a head and T denotes a tail).

(a) What is the probability of heads? What is the probability of tails?

$$P(heads) = \frac{8}{12} = \frac{2}{3}$$

$$P(tails) = \frac{4}{12} = \frac{1}{3}$$

(b) What assumptions did you make in arriving at your answer?

The assumption made of this coin tossing experiment is that it is supposed to be an iid (i.e. independent, identically distributed) process. Because every time we flip a coin:

- the previous result doesn't influence the outcome
- the chances of getting head or tail are identical

Q5

5.(20 pts.) Suppose we have a coin that can remember the outcome of its previous toss. Suppose the first toss has equal probability of being a head or a tail. On each subsequent toss, if the current toss was a head, the next toss will be a tail with probability 1/4 and head with probability 3/4. If the current toss was a tail, the next toss will be a head with probability 1/2 and tail with probability 1/2. What is the probability of observing the sequence H,H,T,H?

$$P(T \mid H) = \frac{1}{4}$$

$$P(H \mid H) = \frac{3}{4}$$

$$P(H \mid T) = \frac{1}{2}$$

$$P(T \mid T) = \frac{1}{2}$$

Thus, the probability of observing the sequence H, H, T, H is:

$$P(H)\, P(H \mid H)\, P(T \mid H)\, P(H \mid T) = \frac{1}{2} \cdot \frac{3}{4} \cdot \frac{1}{4} \cdot \frac{1}{2} = \frac{3}{64}$$