

# Exam 2 Written

Benny Chen

March 31, 2023

## Q1 Support Vector Machine

Suppose we have a hard-margin SVM trained upon a set of data points. If we add new features to these sample points (while retaining all the original features; say, an original data point  $[a, b, c]$  becomes  $[a, b, c, d, e]$  where  $d$  and  $e$  can be any numbers), can the optimal  $\|w_{new}\|$  in the altered SVM be greater than the optimal  $\|w_{old}\|$  in the original SVM?

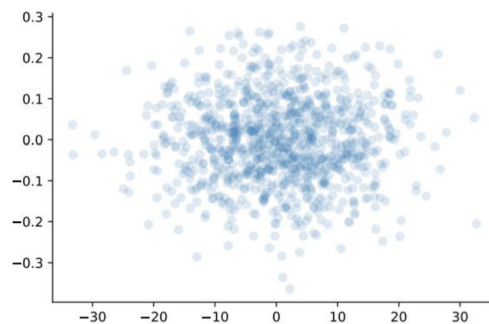
Can it be smaller?

Can it be the same?

**Explain why and most of the points will be for your explanation.**

For a hard-margin SVM, if we add new features to a sample of points, the optimal  $\|w_{new}\|$  can be greater, smaller, and the same as the optimal  $\|w_{old}\|$  in the original SVM. This entirely depends on the new features added. If the new features affect the weight positively or negatively depends on what the new features and weights are. The new weights can change in either direction or stay the same. The only thing that we would know is that there are more datapoints now and are more complex.

## Q2 Principal Component Analysis



The above are 1,000 sample points drawn from a two-dimensional multivariate normal distribution. Which of the following matrices could (without extreme

improbability) be the covariance matrix of the distribution? Pay attention to the numbers on the axes!

**Please provide brief explanation regarding your option and why others are not possible.**

(a)  $\Sigma = \begin{bmatrix} 100 & 0 \\ 0 & 0.01 \end{bmatrix}$

(b)  $\Sigma = \begin{bmatrix} 10 & 0 \\ 0 & 0.1 \end{bmatrix}$

(c)  $\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

(d)  $\Sigma = \begin{bmatrix} -10 & 0 \\ 0 & -0.1 \end{bmatrix}$

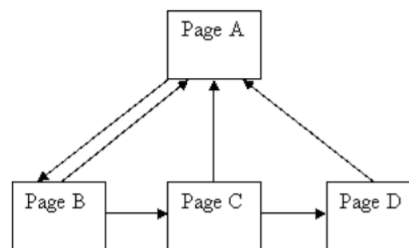
(e) None of the above

From the graph we can see that the data is oriented around the center of the origin and the datapoints are flat and spread out. This shows that there is zero or negligible covariance. This makes our covariance matrix

$$\Sigma = \begin{bmatrix} v_x & 0 \\ 0 & v_y \end{bmatrix} \quad (1)$$

We can now find the variance of the x and y axis to complete the matrix. The variance is the spread of the datapoints. From the spread we can see that it's flat shaped making the variance of the x axis greater than the y axis. This makes c and d not possible. We can also see that it can't be a because the variance is out the range of the data. This leaves us with b which we can conclude as the answer due to it being in the range of the data and the variance of the x axis being greater than the y axis.

### Q3 Page Rank



Regarding the above webpages and their mutual connections, please derive the page ranks (assuming all ranks are 1 initially) after the convergence is observed.

**Warnings: be extremely careful if you have the stochastic weights which do not sum up to exact 1 per webpage. Please make sure the input has the weights per web page that sum up to exact 1! (You will know what I am talking about if you really code above).**

First we make a Stochastic Matrix for the graph which we get:

	A	B	C	D
A	0	0.5	0.5	1
B	1	0	0	0
C	0	0.5	0	0
D	0	0	0.5	0

With all columns summing up to 1. We then plug this into the PageRank formula:

$$r = M * r \quad (2)$$

Where  $r$  is the sum of the columns of the matrix and  $M$  is the Stochastic Matrix. We then solve for  $r$ :

$$r = \begin{bmatrix} 0 & 0.5 & 0.5 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0 \end{bmatrix} * \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \quad (3)$$

This gives us the following:

$$r = \begin{bmatrix} 2 \\ 1 \\ 0.5 \\ 0.5 \end{bmatrix} \quad (4)$$

Which we then keep plugging into the formula until we see a convergence. We then get the following:

$$\begin{bmatrix} 1.45460486 \\ 1.45.446849 \\ 0.72732258 \\ 0.36360407 \end{bmatrix} \quad (5)$$