ML Exam

Syed Abdul Khader

27 December 2022

Problem 1

For $r = 1, 2, 3, 4, 5$ - Find the best rank-r approximation of

0.7820	0.7190	0.5210	0.6280	0.8640	0.5630
1.0750	0.8780	0.6030	0.9290	1.0630	0.8520
1.1250	0.7250	0.4340	1.0860	0.8930	1.0280
0.6600	0.7770	0.6130	0.4290	0.9210	0.3540
0.3600	0.3080	0.2160	0.3020	0.3710	0.2750
1.2270	0.8830	0.5680	1.1300	1.0790	1.0560
0.5520	0.3850	0.2430	0.5160	0.4710	0.4840

Find the $L^2 - norm$ error between the approximation and the original matrix. If we were to design an efficient value of r, which r will we choose and why?

Solution

We know the Singular Value Decomposition for a Matrix A is given by

$$A = U\Sigma V^T$$

To get the values U, Σ and V we use the following

$$A^T A = VU^T UV^T = V\Sigma^2 V^T (U.U^T = I)$$

$$AA^T = UV^T VU^T = U\Sigma^2 U^T (V.V^T = I)$$

$$A^T A = \begin{bmatrix} 5.408 & 4.241 & 2.857 & 4.775 & 5.148 & 4.408 \\ 4.241 & 3.44 & 2.357 & 3.677 & 4.166 & 3.377 \\ 2.857 & 2.357 & 1.628 & 2.454 & 2.851 & 2.247 \\ 4.775 & 3.677 & 2.454 & 4.255 & 4.469 & 3.939 \\ 5.148 & 4.166 & 2.851 & 4.469 & 5.046 & 4.106 \\ 4.408 & 3.377 & 2.247 & 3.939 & 4.106 & 3.65 \end{bmatrix}$$

Characteristic Equation of $A^T A$ is given as $|A^T A - \lambda I| = 0$

$$\begin{vmatrix} 5.408-\lambda & 4.241 & 2.857 & 4.775 & 5.148 & 4.408 \\ 4.241 & 3.44-\lambda & 2.357 & 3.677 & 4.166 & 3.377 \\ 2.857 & 2.357 & 1.628-\lambda & 2.454 & 2.851 & 2.247 \\ 4.775 & 3.677 & 2.454 & 4.255-\lambda & 4.469 & 3.939 \\ 5.148 & 4.166 & 2.851 & 4.469 & 5.046-\lambda & 4.106 \\ 4.408 & 3.377 & 2.247 & 3.939 & 4.106 & 3.65-\lambda \end{vmatrix} = 0$$

The Eigen Values of the above system are

$[22.99959, 0.427182, 0.00000079133, 0.0000003815, 0.0000001551, 0.0000001101855]$

The corresponding eigen vectors for the above eigen values is transposed to get the final V^T matrix.

$$V^T = \begin{bmatrix} -0.484 & -0.383 & -0.259 & -0.426 & -0.465 & -0.393 \\ -0.161 & 0.389 & 0.44 & -0.447 & 0.429 & -0.496 \\ -0.626 & -0.384 & -0.016 & 0.305 & 0.593 & 0.126 \\ -0.039 & 0.432 & -0.839 & -0.147 & 0.293 & -0.007 \\ -0.498 & 0.575 & 0.096 & 0.501 & -0.39 & -0.091 \\ 0.312 & -0.192 & -0.159 & 0.504 & 0.102 & -0.759 \end{bmatrix}$$

The Σ Matrix is a diagonal matrix, with the values having square root of the eigen values.

$$\Sigma = \begin{bmatrix} 4.79578 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.65359 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.00088 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.00061 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.00039 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.00033 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Now for AA^T,

$$AA^T = \begin{bmatrix} 2.857 & 3.767 & 3.659 & 2.658 & 1.280 & 4.126 & 1.838 \\ 3.767 & 5.009 & 4.941 & 3.440 & 1.696 & 5.533 & 2.470 \\ 3.659 & 4.941 & 5.013 & 3.224 & 1.664 & 5.543 & 2.484 \\ 2.658 & 3.440 & 3.224 & 2.572 & 1.177 & 3.696 & 1.638 \\ 1.280 & 1.696 & 1.664 & 1.177 & 0.575 & 1.868 & 0.833 \\ 4.126 & 5.533 & 5.543 & 3.696 & 1.868 & 6.164 & 2.757 \\ 1.838 & 2.470 & 2.484 & 1.638 & 0.833 & 2.757 & 1.234 \end{bmatrix}$$

Characteristic Equation of AA^T is given as $|AA^T - \lambda I| = 0$

The Eigen Values are same as we got in the above case, and the corresponding Eigen Vector forms the U matrix.

$$U = \begin{bmatrix} -0.35 & 0.298 & 0.24 & -0.177 & -0.077 & -0.734 & -0.395 \\ -0.467 & 0.082 & 0.421 & 0.676 & 0.288 & 0.077 & 0.229 \\ -0.462 & -0.488 & -0.362 & 0.139 & 0.018 & 0.192 & -0.601 \\ -0.318 & 0.756 & -0.26 & -0.17 & 0.004 & 0.469 & -0.102 \\ -0.158 & 0.069 & -0.711 & 0.297 & -0.162 & -0.412 & 0.425 \\ -0.517 & -0.258 & 0.215 & -0.391 & -0.525 & 0.138 & 0.415 \\ -0.231 & -0.153 & -0.126 & -0.471 & 0.78 & -0.101 & 0.259 \end{bmatrix}$$

For Rank 1 Approximation we take only the first eigen value in the Σ matrix and get the new A_1 which is Rank 1 approximation of A

$$\Sigma = \begin{bmatrix} 4.79578 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$A_1 = U\Sigma V^T$$

$$A_1 = \begin{bmatrix} 0.813 & 0.643 & 0.435 & 0.715 & 0.78 & 0.659 \\ 1.084 & 0.857 & 0.58 & 0.953 & 1.04 & 0.878 \\ 1.074 & 0.849 & 0.574 & 0.944 & 1.03 & 0.87 \\ 0.739 & 0.585 & 0.395 & 0.65 & 0.709 & 0.599 \\ 0.367 & 0.29 & 0.196 & 0.322 & 0.352 & 0.297 \\ 1.2 & 0.949 & 0.642 & 1.055 & 1.151 & 0.972 \\ 0.536 & 0.424 & 0.287 & 0.471 & 0.514 & 0.434 \end{bmatrix}$$

$L^2 - norm$ between original A and Rank-1 approximation A_1 is given by $\sqrt{\sum_i n \sum_j n |a_{i,j}|^2}$

$$|A - A_1|_2 = 0.6535$$

Similarly Rank-2 Approximation is:

$$\Sigma = \begin{bmatrix} 4.79578 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.65359 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$A_2 = U\Sigma V^T$$

$$A_2 = \begin{bmatrix} 0.782 & 0.719 & 0.521 & 0.628 & 0.864 & 0.563 \\ 1.075 & 0.878 & 0.603 & 0.929 & 1.063 & 0.852 \\ 1.125 & 0.725 & 0.434 & 1.086 & 0.893 & 1.028 \\ 0.66 & 0.777 & 0.613 & 0.429 & 0.921 & 0.354 \\ 0.36 & 0.308 & 0.216 & 0.302 & 0.371 & 0.275 \\ 1.227 & 0.883 & 0.568 & 1.13 & 1.079 & 1.056 \\ 0.552 & 0.385 & 0.243 & 0.516 & 0.471 & 0.484 \end{bmatrix}$$

$$|A - A_2|_2 = 0.001199$$

Similarly Rank-3 Approximation is:

$$\Sigma = \begin{bmatrix} 4.79578 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.65359 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.00088 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$A_3 = U \Sigma V^T$$

$$A_3 = \begin{bmatrix} 0.782 & 0.719 & 0.521 & 0.628 & 0.864 & 0.563 \\ 1.075 & 0.878 & 0.603 & 0.929 & 1.063 & 0.852 \\ 1.125 & 0.725 & 0.434 & 1.086 & 0.893 & 1.028 \\ 0.66 & 0.777 & 0.613 & 0.429 & 0.921 & 0.354 \\ 0.36 & 0.308 & 0.216 & 0.302 & 0.371 & 0.275 \\ 1.227 & 0.883 & 0.568 & 1.13 & 1.079 & 1.056 \\ 0.552 & 0.385 & 0.243 & 0.516 & 0.471 & 0.484 \end{bmatrix}$$

$$|A - A_3|_2 = 0.000804$$

Similarly Rank-4 Approximation is:

$$\Sigma = \begin{bmatrix} 4.79578 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.65359 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.00088 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.00061 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$A_4 = U \Sigma V^T$$

$$A_4 = \begin{bmatrix} 0.782 & 0.719 & 0.521 & 0.628 & 0.864 & 0.563 \\ 1.075 & 0.878 & 0.603 & 0.929 & 1.063 & 0.852 \\ 1.125 & 0.725 & 0.434 & 1.086 & 0.893 & 1.028 \\ 0.66 & 0.777 & 0.613 & 0.429 & 0.921 & 0.354 \\ 0.36 & 0.308 & 0.216 & 0.302 & 0.371 & 0.275 \\ 1.227 & 0.883 & 0.568 & 1.13 & 1.079 & 1.056 \\ 0.552 & 0.385 & 0.243 & 0.516 & 0.471 & 0.484 \end{bmatrix}$$

$$|A - A_4|_2 = 0.000515$$

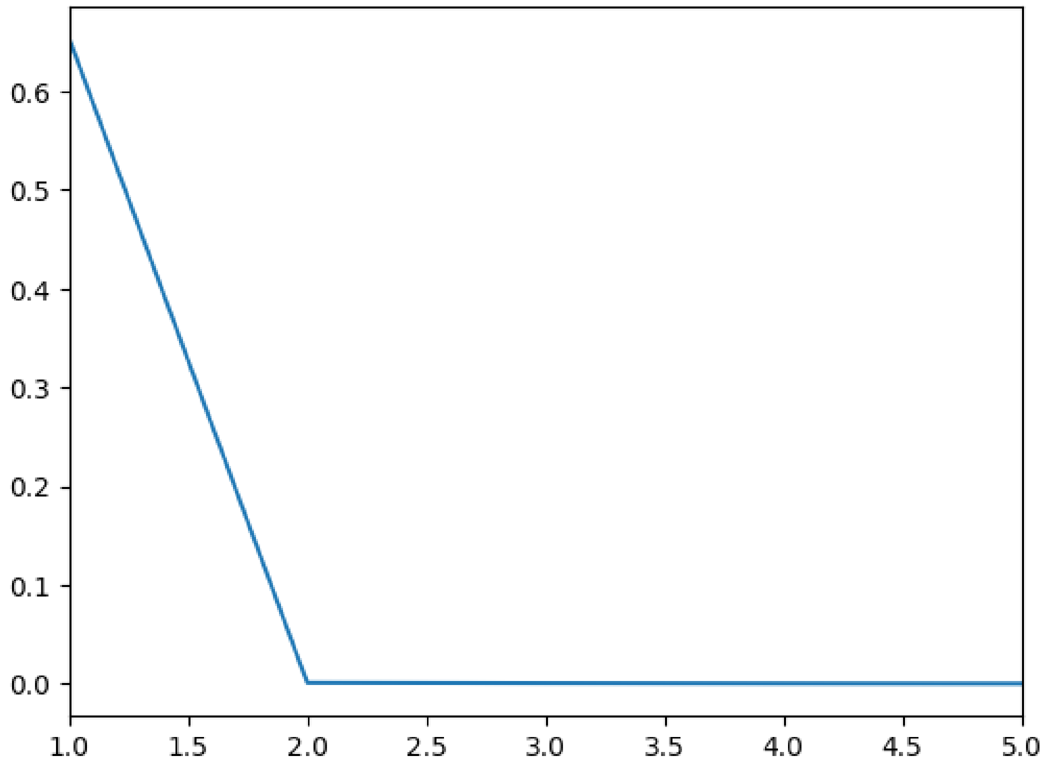
Similarly Rank-5 Approximation is:

$$\Sigma = \begin{bmatrix} 4.79578 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.65359 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.00088 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.00061 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.00039 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$A_5 = U\Sigma V^T$$

$$A_5 = \begin{bmatrix} 0.782 & 0.719 & 0.521 & 0.628 & 0.864 & 0.563 \\ 1.075 & 0.878 & 0.603 & 0.929 & 1.063 & 0.852 \\ 1.125 & 0.725 & 0.434 & 1.086 & 0.893 & 1.028 \\ 0.66 & 0.777 & 0.613 & 0.429 & 0.921 & 0.354 \\ 0.36 & 0.308 & 0.216 & 0.302 & 0.371 & 0.275 \\ 1.227 & 0.883 & 0.568 & 1.13 & 1.079 & 1.056 \\ 0.552 & 0.385 & 0.243 & 0.516 & 0.471 & 0.484 \end{bmatrix}$$

$$|A - A_5|_2 = 0.000331$$



Here we can see that, when $r = 1$, the $L^2 - norm$ is high, whereas as $r = 2$, the error decreases drastically and doesn't reduce much with increase in r. Hence the efficient rank is $r = 2$

Problem 2

Solve the Ridge Regression for the following inputs with $= 0.1$. First 6 columns are the features and the last column is the output.

4.4300	2.9200	3.6800	3.9300	3.4200	4.0500	3.3000
1.5800	1.9300	1.7600	1.7000	1.8200	1.6700	1.8400
0.4400	0.5500	0.4900	0.4800	0.5100	0.4700	0.5200
7.4400	5.4200	6.4300	6.7600	6.0900	6.9300	5.9300
4.7400	6.3400	5.5400	5.2700	5.8100	5.1400	5.9400
6.1500	8.1200	7.1400	6.8100	7.4600	6.6500	7.6300
3.1600	3.7400	3.4500	3.3500	3.5400	3.3000	3.5900
7.1600	4.9400	6.0500	6.4200	5.6800	6.6100	5.5000
2.8200	7.4500	5.1300	4.3600	5.9100	3.9700	6.2900
6.1900	6.3800	6.2900	6.2500	6.3200	6.2400	6.3400

The Training Features are the first 6 columns

$$X = \begin{bmatrix} 4.4300 & 2.9200 & 3.6800 & 3.9300 & 3.4200 & 4.0500 \\ 1.5800 & 1.9300 & 1.7600 & 1.7000 & 1.8200 & 1.6700 \\ 0.4400 & 0.5500 & 0.4900 & 0.4800 & 0.5100 & 0.4700 \\ 7.4400 & 5.4200 & 6.4300 & 6.7600 & 6.0900 & 6.9300 \\ 4.7400 & 6.3400 & 5.5400 & 5.2700 & 5.8100 & 5.1400 \\ 6.1500 & 8.1200 & 7.1400 & 6.8100 & 7.4600 & 6.6500 \\ 3.1600 & 3.7400 & 3.4500 & 3.3500 & 3.5400 & 3.3000 \\ 7.1600 & 4.9400 & 6.0500 & 6.4200 & 5.6800 & 6.6100 \\ 2.8200 & 7.4500 & 5.1300 & 4.3600 & 5.9100 & 3.9700 \\ 6.1900 & 6.3800 & 6.2900 & 6.2500 & 6.3200 & 6.2400 \end{bmatrix}$$

$$Y = \begin{bmatrix} 3.3000 \\ 1.8400 \\ 0.5200 \\ 5.9300 \\ 5.9400 \\ 7.6300 \\ 3.5900 \\ 5.5000 \\ 6.2900 \\ 6.3400 \end{bmatrix}$$

To find the weights using the closed form where the weights $\beta = (x^T x + \lambda I)^{-1} x^T y$. We get the weights of the model.

$$X^T X = \begin{bmatrix} 245.4783 & 244.2314 & 244.9303 & 244.9987 & 244.6208 & 245.1838 \\ 244.2314 & 282.6583 & 263.5016 & 256.9696 & 269.8478 & 253.8369 \\ 244.9303 & 263.5016 & 254.2822 & 251.0534 & 257.2973 & 249.5812 \\ 244.9987 & 256.9696 & 251.0534 & 248.9229 & 252.9613 & 248.0026 \\ 244.6208 & 269.8478 & 257.2973 & 252.9613 & 261.4292 & 250.9325 \\ 245.1838 & 253.8369 & 249.5812 & 248.0026 & 250.9325 & 247.3599 \end{bmatrix}$$

$$X^T X + \lambda I = \begin{bmatrix} 245.4783 + \lambda & 244.2314 & 244.9303 & 244.9987 & 244.6208 & 245.1838 \\ 244.2314 & 282.6583 + \lambda & 263.5016 & 256.9696 & 269.8478 & 253.8369 \\ 244.9303 & 263.5016 & 254.2822 + \lambda & 251.0534 & 257.2973 & 249.5812 \\ 244.9987 & 256.9696 & 251.0534 & 248.9229 + \lambda & 252.9613 & 248.0026 \\ 244.6208 & 269.8478 & 257.2973 & 252.9613 & 261.4292 + \lambda & 250.9325 \\ 245.1838 & 253.8369 & 249.5812 & 248.0026 & 250.9325 & 247.3599 + \lambda \end{bmatrix}$$

The Determinant of $X^T X$ is 3.58, which is non-zero. Hence it can be inverted.

Here $\lambda = 0.1$

$$(X^TX+\lambda I)^{-1} = \begin{bmatrix} 4.8894 & 2.3968 & -1.3569 & -2.6011 & -0.0986 & -3.2274 \\ 2.3968 & 3.5241 & -2.0308 & -0.5583 & -3.5181 & 0.1856 \\ -1.3569 & -2.0308 & 8.2985 & -1.5819 & -1.8062 & -1.5249 \\ -2.6011 & -0.5583 & -1.5819 & 8.0736 & -1.2343 & -2.0940 \\ -0.0986 & -3.5181 & -1.8062 & -1.2343 & 7.6092 & -0.9505 \\ -3.2274 & 0.1856 & -1.5249 & 2.0940 & -0.9505 & 7.6119 \end{bmatrix}$$

$$(X^TX+\lambda I)^{-1} * x^T =$$

$$\begin{bmatrix} 0.034 & -0.029 & -0.011 & 0.091 & -0.017 & -0.07 & 0.019 & 0.045 & -0.054 & 0.001 \\ -0.04 & -0.028 & 0.023 & -0.039 & 0.024 & 0.043 & 0.036 & -0.057 & 0.106 & -0.02 \\ 0.027 & 0.018 & -0.045 & -0.006 & -0.003 & 0.027 & 0.02 & -0.038 & -0.011 & 0.023 \\ 0.051 & 0.01 & 0.035 & -0.003 & -0.021 & 0.021 & 0.001 & 0.026 & -0.018 & -0.023 \\ -0.035 & 0.038 & -0.022 & -0.009 & 0.039 & -0.034 & -0.037 & -0.002 & 0.06 & 0.026 \\ -0.02 & -0.003 & 0.022 & -0.005 & -0.002 & 0.038 & -0.026 & 0.054 & -0.07 & 0.018 \end{bmatrix}$$

$$\beta = \begin{bmatrix} -0.05125453 \\ 0.42471989 \\ 0.18631824 \\ 0.10696734 \\ 0.26629343 \\ 0.06744087 \end{bmatrix}$$