CS 189: Introduction to Machine Learning - Discussion 12

1. SVD Warmup

Find the SVD of $X = \begin{bmatrix} 4 & 4 \\ 3 & -3 \end{bmatrix}$.

Solution: First, let's calculate XX^T and X^TX.

$$XX^T = \begin{bmatrix} 32 & 0 \\ 0 & 18 \end{bmatrix}$$

$$X^TX = \begin{bmatrix} 25 & 7 \\ 7 & 25 \end{bmatrix}$$

We find the eigenvectors of X^TX to be $\begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$ and $\begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$ with corresponding eigenvalues 32 and 18.

We use these results to calculate the left singular vectors.

$$Xv_i = \sigma_i u_i$$

$$u_i = Xv_i/\sigma_i$$

$$u_1 = \begin{bmatrix} 4 & 4 \\ 3 & -3 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} / \sqrt{32} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$u_1 = \begin{bmatrix} 4 & 4 \\ 3 & -3 \end{bmatrix} \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} / \sqrt{18} = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$$

Thus, the SVD of X is $X = U\Sigma V^T = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \sqrt{32} & 0 \\ 0 & \sqrt{18} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$

2. Derivation of PCA

In this question we will derive PCA. PCA aims to find the direction of maximum variance among a dataset. You want the line such that projecting your data onto this line will retain the maximum amount of information. Thus, the optimization problem is

$$\max_{u:\|u\|_2=1} \frac{1}{n} \sum_{i=1}^{n} \left(u^T x_i - u^T \hat{x}\right)^2$$

where n is the number of data points and $\hat{x}$ is the sample average of the data points.

(a) Show that this optimization problem can be massaged into this format

$$\max_{u:\|u\|_2=1} u^T \Sigma u$$

where $\Sigma = \frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{x})(x_i - \hat{x})^T$.

Solution:
We can massage the objective function (left's call if $f_0(u)$ in this way:

$$\begin{aligned}
f_0(u) &= \frac{1}{n} \sum_{i=1}^{n} \left(u^T x_i - u^T \hat{x}\right)^2 \\
&= \frac{1}{n} \sum_{i=1}^{n} \left((x_i - \hat{x})^T u\right)^2 \\
&= \frac{1}{n} \sum_{i=1}^{n} (u^T (x_i - \hat{x}))((x_i - \hat{x})^T u) \\
&= u^T \left(\frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{x})(x_i - \hat{x})^T \right) u \\
&= u^T \Sigma u
\end{aligned}$$

(b) Show that the maximizer for this problem is equal to v_1, where v_1 is the eigenvector corresponding to the largest eigenvalue λ_1. Also show that optimal value of this problem is equal to λ_1.

Solution:
We start by invoking the spectral decomposition of $\Sigma = V \Lambda V^T$, which is a symmetric positive semi-definite matrix.

$$\begin{aligned}\max_{u:\|u\|_2=1} u^T\Sigma u &= \max_{u:\|u\|_2=1} u^T V\Lambda V^T u \\ &= \max_{u:\|u\|_2=1} (V^T u)^T \Lambda V^T u\end{aligned}$$

Here is an aside: note through this one line proof that left-multiplying a vector by an orthogonal (or rotation) matrix preserves the length of the vector:

$$\|V^T u\|_2 = \sqrt{(V^T u)^T (V^T u)} = \sqrt{u^T V V^T u} = \sqrt{u^T u} = \|u\|_2$$

I define a new variable $z = V^T u$, and maximize over this variable. Note that because V is invertible, there is a one to one mapping between u and z. Also note that the constraint is the same because the length of the vector u does not change when multiplied by an orthogonal matrix.

$$\max_{z:\|z\|_2=1} z^T \Lambda z = \max_z \sum_{i=1}^d \lambda_i z_i^2 \; : \; \sum_{i=1}^d z_i^2 = 1$$

From this new formulation, it is obvious to see that we can maximize this by throwing all of our eggs into one basket and setting $z_i^* = 1$ if i is the index of the largest eigenvalue, and $z_i^* = 0$ otherwise. Thus,

$$z^* = V^T u^* \implies u^* = V z^* = v_1$$

where v_1 is the "principle" eigenvector, and corresponds to λ_1. Plugging this into the objective function, we see that the optimal value is λ_1.

3. Deriving the second principal component

(a) Let $J(\mathbf{v_2}, \mathbf{z_2}) = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x_i} - z_{i1}\mathbf{v_1} - z_{i2}\mathbf{v_2})^T(\mathbf{x_i} - z_{i1}\mathbf{v_1} - z_{i2}\mathbf{v_2})$ given the constraints $\mathbf{v_1^T v_2} = 0$ and $\mathbf{v_2^T v_2} = 1$. Show that $\frac{\partial J}{\partial \mathbf{z_2}} = 0$ yields $z_{i2} = \mathbf{v_2}^T\mathbf{x_i}$.

(b) We have shown that $z_{i2} = \mathbf{v_2}^T\mathbf{x_i}$ so that the second principal encoding is gotten by projecting onto the second principal direction. Show that the value of $\mathbf{v_2}$ that minimizes J is given by the eigenvector of $\mathbf{C} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x_i x_i^T})$ with the second largest eigenvalue. Assumed we have already proved the v_1 is the eigenvector of C with the largest eigenvalue.

Solution: (a) We have

$$J(\mathbf{v_2}, \mathbf{z_2}) = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x_i^T x_i} - z_{i1}\mathbf{x_i^T v_1} - z_{i2}\mathbf{x_i^T v_2} - z_{i1}\mathbf{v_1^T x_i} + z_{i1}^2\mathbf{v_1^T v_1} + \tag{1}$$

$$z_{i1}z_{i2}\mathbf{v_1^T v_2} - z_{i2}\mathbf{v_2^T x_i} + z_{i1}z_{i2}\mathbf{v_2^T v_1} + z_{i2}^2\mathbf{v_2^T v_2}) \tag{2}$$

Take derivative respect to $\mathbf{z_2}$, we have

$$\frac{\partial J}{\partial z_{i2}} = \frac{1}{n}(-\mathbf{x_i^T v_2} + z_{i1}\mathbf{v_1^T v_2} - \mathbf{v_2^T x_i} + z_{i1}\mathbf{v_2^T v_1} + 2z_{i2}\mathbf{v_2^T v_2}) = \frac{1}{n}(-2\mathbf{x_i^T v_2} + 2z_{i2}\mathbf{v_2^T v_2})$$

Set the derivative to 0 and we have

$$z_{i2}\mathbf{v_2^T v_2} = \mathbf{x_i^T v_2}$$

Since $\mathbf{v_2^T v_2} = 1$, we have $z_{i2} = \mathbf{x_i^T v_2}$

(b) Plug in z_{i2} into $J(\mathbf{v_2}, \mathbf{z_2})$, we have

$$J(\mathbf{v_2}) = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x_i^T x_i} - z_{i1}\mathbf{x_i^T v_1} - z_{i2}\mathbf{x_i^T v_2} - z_{i1}\mathbf{v_1^T x_i} + z_{i1}^2\mathbf{v_1^T v_1} - z_{i2}\mathbf{v_2^T x_i} + z_{i2}^2\mathbf{v_2^T v_2})$$

$$= \frac{1}{n}\sum_{i=1}^{n}(const - 2z_{i2}\mathbf{x_i^T v_2} + z_{i2}^2) = \frac{1}{n}\sum_{i=1}^{n}(-2\mathbf{v_2^T x_i x_i^T v_2} + \mathbf{v_2^T x_i x_i^T v_2} + const)$$

$$= -\mathbf{v_2^T C v_2} + const$$

In order to minimize J with constraints $\mathbf{v_2^T v_2} = 1$, we have Langrage $L = -\mathbf{v_2^T C v_2} + \lambda(\mathbf{v_2^T v_2} - 1)$ and take derivative of $\mathbf{v_2}$, we have

$$\frac{\partial L}{\partial \mathbf{v_2}} = -2\mathbf{C v_2} + 2\lambda v_2 = 0$$

Then, we have

$$\mathbf{C v_2} = \lambda \mathbf{v_2}$$