BASIC ALGEBRA HW

Ufuk Cem Birbiri

Ex-1

a) Let $y(u) = Su$, so that

$$f(u, y(u)) = u^T \cdot y(u).$$

$$\rightarrow f'(u, y(u)) = \frac{\partial f}{\partial u} + \frac{\partial f}{\partial y} \cdot y'(u)$$

The first term $\frac{\partial f}{\partial u}$ is the derivative of inner product. If we hold $y(u)$ fixed, then the derivative of $u^T \cdot y(u)$ is $y(u)^T$.

Likewise, in the second term, $\frac{\partial f}{\partial y}$ is equal to u^T.

And $y'(u) = \frac{\partial y}{\partial u}$ is the derivative of a matrix times a vector.

So, $y'(u) = \frac{\partial (Su)}{\partial u} = S$.

$$\Rightarrow f'(u, y(u)) = y(u)^T + u^T S$$

Since $y(u) = Su$, we have $y(u)^T = u^T \cdot S^T$

$$f'(u, y(u)) = u^T \cdot S^T + u^T S = u^T (S^T + S)$$

Since S is symmetric, $S^T = S$

$$f'(u, y(u)) = u^T (S^T + S) = u^T \cdot (2S) = 2 u^T S$$

b) $g(U) := U^T U$.

$\nabla g(U) = \nabla(U^T U)$. $\longrightarrow$ Lets say $U = (U_1, U_2, \dots U_n)$

Lets write $g(U) = U^T U = \sum_{i=1}^{N} U_i^2$.

$$\frac{d}{dU_1}(U^T U) = \frac{d}{dU_1}\left(\sum_{i=1}^{N} U_i^2\right) = \frac{d}{dU_1} x_1^2 = 2x_1$$

We can compute the derivative for each term of U.

So $\frac{d}{dU}(U^T U) = 2U$

---

a)

<u>Theorem</u>: Max and Min of $f(U)$ given $\|U\| = 1$

- Let A be a symmetric matrix that defines a quadratic form $f(U) = U^T A U$. Then, under the condition that U is a unit vector, the <u>maximum</u> value of $f(U)$ is the largest eigenvalue λ_{max}, and the <u>minimum</u> value of $f(U)$ is the smallest eigenvalue λ_{min}.

* $\max\{f(U) : \|U\| = 1\} = \lambda_{max}$

* $\min\{f(U) : \|U\| = 1\} = \lambda_{min}$.

In our case, S is a positive definite symmetric matrix

$S = Q \Lambda Q^T$.

- $f(u) = u^T . S . u = u^T Q \Lambda Q^T u$
- If $\|u\| = 1$, then $\lambda_{max} = \max \{ f(u) \}$
  or $\lambda_{min} = \min \{ f(u) \}$.
- we know that $u \in \mathcal{U}$ where $\mathcal{U} := \{ u \in R^D \mid \|u\|_2 = 1 \}$.

So, we can write $f(u) = u^T Q \Lambda Q^T u = u^T S u$

Here, S is a symmetric matrix where $S \in R^{D \times D}$.

Also, Λ is a symmetric matrix where it has eigenvalues of S in the main diogonal in decreasing order.

Let the new form of $f(u) = u^T Q \Lambda Q^T u = z^T \Lambda z$

where $z = Q^T u$.

Norm of the z should be equal to 1 ~~we want to same~~ since $f(u)$ didn't change and only S is swithed by Λ, which is symmetric and $\Lambda \in R^{D \times D}$.

So, $\|Q^T u = z\| = 1$

---

b) If the $\|u\| = 1$, then $\max \{ f(u) : \|u\| = 1 \} = \lambda_{max}$.

So, $\max_{u \in \mathcal{U}} (f(u)) = \max_{u \in \mathcal{U}} (u^T S u) \leq \lambda_{max} = \lambda_1$

where $\lambda_1 > \lambda_2 > \dots > \lambda_D$

Ex-2 | a vector space V of dim. N, and a sub-vector space $W \subset V$ with an orthonormal basis $\{u_1, \dots, u_k\}$.

Orthogonal Projection: $P_W : V \to W$.

a) Show that the projection matrix associated to P_W can be written in the form $M_{P_W} = UU^T$. What're the columns of U?

- If $M_{P_W} = U.U^T$ is a projection matrix, then the colums of U are orthonormal, we have $U^TU = I$.

Clearly, $(UU^T)^T = (U^T)^T.U^T = U.U^T$

- ~~If $u_1, \dots, u_k$ are orthonormal vectors spanning N dimensional subspace~~ ,
~~U is the~~

- The columns of U are u_i's $\to$ orthonormal basis $= \{u_1 \dots u_k\}$

b) $M_{P_W} = U.U^T$ is symetric since:

$(UU^T)^T = (U^T)^TU^T = U.U^T$ since $U^TU = I$

It is idempotents since:

$(UU^T)^2 = (UU^T)(UU^T) = U(U^TU)U^T = UIU^T = U.U^T$
(since $U^TU = I$)

Question-3

a) $y_i = U^T . x_i$

$E[y_1] = E[U^T x_1]$

$E[y_1] = U^T . E[x_1]$ → since Expectation is a linear operator.

$\overline{y} = U^T . \overline{x}$

(5)

Question-3

b) We are given N data points $x_1 \dots x_N$ in R^D.
Let's put these data points into the rows of an $N \times D$ matrix X:

$$X = \begin{bmatrix} - x_1^T - \\ \vdots \\ - x_N^T - \end{bmatrix} \in R^{N \times D}$$

For PCA to work correctly, we will need to subtract off the sample mean from each row, in particular denote by $\bar{x}$ the sample mean:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i ,$$

and for each $i \in [N]$, denote by $\tilde{x}_i$ the i-th data point after subtracting mean:

$$\tilde{x}_i = x_i - \bar{x}.$$

Let denote the resulting mean-centered data matrix:

$$\tilde{X} = \begin{bmatrix} - \tilde{x}_1^T - \\ \vdots \\ - \tilde{x}_N^T - \end{bmatrix} \in R^{N \times D}$$

The $D \times D$ sample covariance matrix S can be written in terms of $\tilde{X}$ as:

$$S = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})(x_i - \bar{x})^T = \frac{1}{N-1} \tilde{X}^T \cdot \tilde{X}.$$

Now, first consider projecting to one dimension, say along a direction v_1 for some unit vector $v_1 \in \mathbb{R}^D$.
For any direction v, the sample variance of the projection data points
$v^T \tilde{x}_1, \dots, v^T \tilde{x}_N$ is given by:

$$\frac{1}{N-1} \sum_{i=1}^{N} (v^T \tilde{x}_i)^2 = \frac{1}{N-1} \|\tilde{X} v\|_2^2 = \frac{1}{N-1} v^T \tilde{X}^T \tilde{X} v = v^T S v$$

- We wish to find $v_1 \in \mathbb{R}^D$ (with $\|v_1\|_2 = 1$) s.t. the sample variance of the projected data points is maximized:

$$v_1 \in \operatorname*{argmax}_{v \in \mathbb{R}^D : \|v\|_2 = 1} v^T S v$$

Here, in our case $v = U$

- It is well know that this maximum is achieved by choosing v_1 to be an eigenvector of S corresponding to the largest eigenvalue of S. Indeed, introducing a Lagrange multiplier λ for the constraint $\|v\|_2^2 = 1$,

$\longrightarrow$

, we obtain the Lagrangian function

$$-v^T S v + \lambda(\|v\|_2^2 - 1);$$

minimizing this over v gives:

$$Sv = \lambda v$$

showing that v must be an eigenvector of S. Moreover, we have that the sample variance of the projected data points is then given by

$$v^T S v = \lambda \|v\|_2^2 = \lambda,$$ the eigenvalue corresponding to this eigenvector. Thus, the variance is maximized by taking λ to be the largest eigenvalue λ_1 of S, and v to be the corresponding (normalized) eigenvector v_1.

According to explanation above,

(a) $U_1, \dots, U_k$'s are colums of U. In the explanation $v = U$.

(b,c,d) Setting the derivative of J:

$$J = -v^T S v + \lambda\left(\|v\|_2^2 - 1\right)$$ wrt v to 0, we get a stationary point $\rightarrow$

$$\frac{\partial J}{\partial v} = (-2Sv) + 2\lambda v = 0$$

$$Sv = \lambda v.$$

This shows that at the stationary point, v must be an eigenvector of S, and λ be an eigenvalue corresponding to the eigenvector v.

- Left-multiplying the equation,

$J = -v^T S v + \lambda (\|v\|_2^2 - 1)$ with v^T, we can see that the maximum variance is equal to the eigenvalue λ.

$$v^T S . v = v^T \lambda v$$

$v^T S v = \lambda v^T v$ → since λ is a scalar

$v^T S v = \lambda$ → since $v^T v = 1$

Again, in this explanation $V = U$ and $u_1, \ldots u_k$ are the colums of $V = U$.