S&DS 365 Homework 6 Solutions

Yale University, Department of Statistics

April 20, 2021

1 PCA

Problem 1:

Part 1)

Since the $\tilde{v}_i$ vectors are orthonormal, they form a basis and any vector u can be written in the components of the basis $u = \sum_{i=1}^{p} \langle u, \tilde{v}_i \rangle \tilde{v}_i$. If we then but u through $\sigma^2 \sum_i \tilde{v}_i \tilde{v}_i^T$ as have

$$
\begin{aligned}
(\sigma^2 \sum_i \tilde{v}_i \tilde{v}_i^T)u &= \sigma^2 \sum_i v_i(v_i^T u) \\
&= \sigma^2 \sum_i \tilde{v}_i(\langle u, \tilde{v}_i \rangle) \\
&= \sigma^2 u
\end{aligned}
$$

since the holds for any vector u the matrix in question must be $\sigma^2 I$.

Part 2)

$$
\begin{aligned}
vv^T + \sigma^2 I &= vv^T + \sigma^2 \sum_{i=1}^{p} \tilde{v}_i \tilde{v}_i^T \\
&= (\|v_1\|_2 \tilde{v_1})((\|v_1\|_2 \tilde{v_1})^T + \sigma^2 \sum_{i=1}^{p} \tilde{v}_i \tilde{v}_i^T \\
&= \|v_1\|_2^2 \tilde{v_1} \tilde{v_1}^T + \sigma^2 \sum_{i=1}^{p} \tilde{v}_i \tilde{v}_i^T
\end{aligned}
$$

Part 3)

Taking the above just move one term from the sum over to the other term,

$$
\begin{aligned}
&(\|v_1\|_2^2 + \sigma^2)\tilde{v_1}\tilde{v_1}^T + \sigma^2 \sum_{i=2}^{p} \tilde{v}_i \tilde{v}_i^T \\
&= \begin{pmatrix} | & \cdots & | \\ v_1 & \cdots & v_p \\ | & \cdots & | \end{pmatrix} \begin{pmatrix} \sigma^2 + \|v\|_2^2 & 0 & \cdots \\ 0 & \sigma^2 & 0 \\ \ddots & \ddots & \ddots \end{pmatrix} \begin{pmatrix} | & \cdots & | \\ v_1 & \cdots & v_p \\ | & \cdots & | \end{pmatrix}^T
\end{aligned}
$$

The so largest top right singular value is $\|v\|_2^2 + \sigma^2$ and the rest are σ^2.

Problem 2:

This follows pretty straight forward from above, the top right singular value is $\|v\|_2^2 + \sigma^2$ and it's corresponding vector must be $v_1 = \frac{v}{\|v\|}$ by the structure of the matrix.

Problem 3:

$$\begin{aligned}
E[x_i] &= 0 \\
E[x_i x_i^T] &= E[(vy_i + w_i)(vy_i + w_i)^T] \\
&= E[y_i^2 vv^T] + 2y_i E[v^T w_i] + E[w_i w_i^T] \\
&= y_i^2 vv^T + \sigma^2 I + 2y_i v^T E[w_i] \\
&= y_i^2 vv^T + \sigma^2 I + 0
\end{aligned}$$

Standard Multivariate normal random vectors are rotationally invariant, meaning if you change the basis to be any set of orthonormal directions you still have a standard normal in that new basis. Therefore, consider our standard normal in the basis of $\tilde{v}_1, \cdots, \tilde{v}_p$. In all the directions of $\tilde{v}_j, j \neq 1$ we only have a standard Gaussian random variable coming from w. However, in the direction of $\tilde{v}_1$ we have a standard Gaussian noise random variable from w as well as a Gaussian with variance $\|v\|^2$ coming from $y_i v$ thus there is more variance in this direction than others since the variance compounds in this direction.

2 SVD

Problem 4

Here we basically show that normalizing along the way cancels and we can equivalently normalize at the end.

Part 1)

$$\begin{aligned}
u &= \frac{Xv}{\|Xv\|_2} \\
v' &= \frac{X^T u}{\|X^T u\|_2} \\
&= \frac{X^T \frac{Xv}{\|Xv\|_2}}{\|X^T \frac{Xv}{\|Xv\|_2}\|_2} \\
&= \frac{X^T Xv}{\|X^T Xv\|_2} \\
u' &= \frac{Xv'}{\|Xv'\|_2} \\
&= \frac{X \frac{X^T Xv}{\|X^T Xv\|_2}}{\|X \frac{X^T Xv}{\|X^T Xv\|_2}\|_2} \\
&= \frac{XX^T Xv}{\|XX^T Xv\|_2}
\end{aligned}$$

Part 2)

Recall $u_1^T u_2 = 0$ and same for v, as well then are normal so $u_1^T u_1 = 1$ and same for v.

$$\begin{aligned}
u &= \frac{Xv}{\|Xv\|_2} \\
&= \frac{(\sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T)v)}{\|(\sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T)v)\|_2} \\
&= \frac{\alpha_1 \sigma_1 u_1 + \alpha_2 \sigma_2 u_2}{\|\alpha_1 \sigma_1 u_1 + \alpha_2 \sigma_2 u_2\|_2} \\
&= \frac{\alpha_1 \sigma_1 u_1 + \alpha_2 \sigma_2 u_2}{\sqrt{\alpha_1^2 \sigma_1^2 + \alpha_2^2 \sigma_2^2}} \\
&= \frac{u_1 + \frac{\alpha_2 \sigma_2}{\alpha_1 \sigma_1} u_2}{\sqrt{1 + \frac{\alpha_2^2 \sigma_2^2}{\alpha_1^2 \sigma_1^2}}}
\end{aligned}$$

$$\begin{aligned}
XX^T Xv &= (\sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T)(\sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T)^T (\sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T)v \\
&= (\sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T)(\sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T)^T (\alpha_1 \sigma_1 u_1 + \alpha_2 \sigma_2 u_2) \\
&= (\sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T)(\alpha_1 \sigma_1^2 v_1 + \alpha_2 \sigma_2^2 v_2) \\
&= \alpha_1 \sigma_1^3 u_1 + \alpha_2 \sigma_2^3 u_2
\end{aligned}$$

therefore normalize this,

$$\begin{aligned}
u' &= \frac{\alpha_1 \sigma_1^3 u_1 + \alpha_2 \sigma_2^3 u_2}{\|\alpha_1 \sigma_1^3 u_1 + \alpha_2 \sigma_2^3 u_2\|_2} \\
&= \frac{\alpha_1 \sigma_1^3 u_1 + \alpha_2 \sigma_2^3 u_2}{\sqrt{\alpha_1^2 \sigma_1^6 + \alpha_2^2 \sigma_2^6}} \\
&= \frac{u_1 + \frac{\alpha_2 \sigma_2^3}{\alpha_1 \sigma_1^3} u_2}{\sqrt{1 + \frac{\alpha_2^2 \sigma_2^6}{\alpha_1^2 \sigma_1^6}}}
\end{aligned}$$

Problem 3

Check the inner product of u and u' with u_1 and see which is larger

$$\begin{aligned}
\langle u, u_1 \rangle &= \frac{1}{\sqrt{1 + \left(\frac{\alpha_2 \sigma_2}{\alpha_1 \sigma_1}\right)^2}} \\
\langle u', u_1 \rangle &= \frac{1}{\sqrt{1 + \left(\frac{\alpha_2 \sigma_2}{\alpha_1 \sigma_1}\right)^2 \left(\frac{\sigma_2}{\sigma_1}\right)^4}}
\end{aligned}$$

Clearly, we see the bottom of the u' fraction is being made smaller since $\frac{\sigma_2}{\sigma_1} < 1$ so we have

$$\begin{aligned} 1 + \left(\frac{\alpha_2\sigma_2}{\alpha_1\sigma_1}\right)^2 \left(\frac{\sigma_2}{\sigma_1}\right)^4 &< 1 + \left(\frac{\alpha_2\sigma_2}{\alpha_1\sigma_1}\right)^2 \\ \sqrt{1 + \left(\frac{\alpha_2\sigma_2}{\alpha_1\sigma_1}\right)^2 \left(\frac{\sigma_2}{\sigma_1}\right)^4} &< \sqrt{1 + \left(\frac{\alpha_2\sigma_2}{\alpha_1\sigma_1}\right)^2} \\ \frac{1}{\sqrt{1 + \left(\frac{\alpha_2\sigma_2}{\alpha_1\sigma_1}\right)^2 \left(\frac{\sigma_2}{\sigma_1}\right)^4}} &> \frac{1}{\sqrt{1 + \left(\frac{\alpha_2\sigma_2}{\alpha_1\sigma_1}\right)^2}} \end{aligned}$$

thus u' is closer to u_1 than u originally was.

Extra Credit

The singular vectors $u_1, \cdots, u_d$ and $v_1, \cdots, v_d$ are orthonormal, therefore take any unit vector q and write it in the components of v, $q = \sum_{i=1}^{d} \alpha_i v_i$ and some vector r in the components of u, $r = \sum_{i=1^d} \beta_i u_i$ where $\sum_{i=1}^{d} \alpha_i^2 \leq 1, \sum_{i=1}^{d} \beta_i^2 \leq 1$. Then we have

$$r^T X v = \sum_{i=1}^{d} \alpha_i \beta_i \sigma_i$$

σ_1 is the largest singular value, therefore this sum is maximized when we put all out mass there and we have $\alpha_1 = \beta_1 = 0$ and this is our maximizer.