HW 4 P1

We can demonstrate this by proving that PCA for reduction to k dimensions takes the eigenvectors corresponding to highest k/we eigenvalues of Σ and then setting $k=2$.

So our problem is now deriving PCA for $k=2$.

Goal: ~~Get P_{kxd} such that $P_{kxd}\, x_{dx1}$ gives x'_{kx1}~~, and we get 'best' P.

Get P_{kxd} s.t. $x'_{kx1} = P_{kxd}\, x_{dx1}$, and we get 'best' P

or get P s.t. reconstruction loss is minimum or

$$\text{if } \hat{x} = P^{-1} x'$$

$$\hat{x} = P^{-1} P x = P^T P x$$ (since P is a set of directions over which we project vectors, it is orthogonal)

$$\text{minimize} \quad \| x - \hat{x} \|$$

Basically: $$\min \| X - (Xp)p^T \| \quad \text{s.t. } p^T p = 1.$$

Looking at PCA as covariance maximisation

If we look at PCA as best set of directions such that variability of data is maximised in lower dimensions,

$$X' = Xp \quad (p: \text{direction})$$

$$\max \| X' \|^2$$

$$= \max \| Xp \|^2 \quad \text{st } p^T p = 1$$

$$= \max \operatorname{tr}\left((Xp)^T (Xp) \right) \quad \text{st } p^T p = 1$$

$$= \max p^T X^T X p \quad \text{st } p^T p = 1$$

$$= \max p^T S p \quad \text{st } p^T p = 1$$

$S = X^T X$ is scatter matrix.

Equivalence of the 2:

$$\min \| X - (Xp)p^T \|^2 \quad \text{st } p^T p = 1$$

$$= \min \operatorname{tr}\left((X - (Xp)p^T)(X - (Xp)p^T)^T \right) \quad \text{st } p^T p = 1$$

$$= \min \operatorname{tr}\left((X - (Xp)p^T)(X^T - pp^T X^T) \right) \quad \text{''} \ldots$$

$$= \min \operatorname{tr}\left(XX^T - 2Xpp^T X^T + Xp(p^T p)p^T X^T \right) \quad \ldots$$

$$= \min \operatorname{tr}(XX^T) - \operatorname{tr}(Xpp^T X^T) \quad \text{st } p^T p = 1$$

$$= \min - \operatorname{tr}(Xpp^T X^T) \quad \text{st } p^T p = 1$$

$$= \min - \operatorname{tr}(pX^T X p) \quad \text{st} \ldots$$

$$= \max p S p \quad \text{st } p^T p = 1.$$

Now, these 2 are equivalent. proved.

Deriving algorithm

PCA: max $p^T S p$ st $p^T p = 1$.

We use Lagrangian multipliers to convert this to a constrained optimisation problem

$$\max \; p^T S p - \lambda (p^T p - 1) \quad \text{st. } \lambda > 0.$$

so objective: $L(p, \lambda) = p^T S p - \lambda (p^T p - 1)$, maximise L.

$\frac{\partial}{\partial L}$, set to 0: $p^T p - 1 = 0$: $p^T p = 1$.

$\frac{\partial}{\partial p}$, set to 0: $2Sp - 2\lambda p = 0$. $Sp = \lambda p$.

now, $L(p, \lambda) = p^T (Sp) - \lambda (p^T p - 1)$

$= p^T \lambda p - \lambda (1-1) = \lambda p^T p = \lambda$.

So optimisation is
$(\max \lambda)$ st. $Sp = \lambda p$

= eigen vector p of scatter matrix S corresponding to max eigenvalue λ. for a single direction.

For n directions, we get eigen vectors corresponding to each of the eigenvalues sorted in descending order.

So we get from this that if $k=2$, the second largest eigenvector-value pair is picked (proved).