Optimization-Based Data Analysis – Brett Bernstein

Recitation 2

1. For a fixed matrix $A \in \mathbb{R}^{m\times n}$ arrange the following values in order: $\|A\|_F$, $\|A\|$, $\|A\|_*$.

   Solution. We have $\|A\|_* \geq \|A\|_F \geq \|A\|$ by the corresponding inequalities on the vectors of singular values.

2. Let $A \in \mathbb{R}^{m\times m}$ with SVD $A = USV^T$.

   (a) Assuming A is invertible, give the SVD for A^{-1}.

   (b) Give the SVD for A^T.

   (c) What is the relationship between the SVD of a symmetric matrix and the diagonal factorization given by the spectral theorem?

   (d) Give the SVD for $A = \mathcal{P}_{\mathcal{S}}$, the orthogonal projection onto the subspace $\mathcal{S} \subseteq \mathbb{R}^m$.

   Solution.

   (a) $A^{-1} = VS^{-1}U^T$

   (b) $A^T = VSU^T$

   (c) By the spectral theorem, we have $A = VSV^T$. If A is positive semidefinite then this is also an SVD. Otherwise, we can compute an SVD by $A = \widetilde{V}|S|V^T$ where $\widetilde{V}_{:,i} = V_{:,i}$ if $S_{ii} \geq 0$ and $\widetilde{V}_{:,i} = -V_{:,i}$ if $S_{ii} < 0$. Here $|S|$ is obtained from S by taking the absolute value of each entry.

   (d) Let U be a matrix whose columns form an orthonormal basis for $\mathcal{S}$. Then $A = UU^T = UIU^T$. Note here that I is a $k \times k$ matrix where k is the dimension of $\mathcal{S}$.

3. Suppose you are given a dataset $\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^m$ as columns of a matrix $X \in \mathbb{R}^{m\times n}$. Your goal is to reduce the dimensionality of the data using PCA.

   (a) Suppose you want the resulting reduced vectors to be in $\mathbb{R}^k$. Explain how to obtain this using PCA.

   (b) How do you determine an appropriate k?

   (c) How do you determine the amount of sample variance in the first principal direction?

   Solution.

   (a) Define $\bar{x} := \frac{1}{n}\sum_{k=1}^n \vec{x}_k$. Let $\widetilde{X}$ denote the centered data matrix with $\widetilde{X}_{:,k} := \vec{x}_k - \bar{x}$. Compute the SVD $\widetilde{X} = USV^T$. The reduced vectors are computed by $U_{:,1:k}^T\widetilde{X}$.

(b) Plot the elements of S in decreasing order, and choose an index k that accounts for all the "large" singular values. Often people look for an "elbow" in the curve.

(c) It is given by $S_{11}^2/(n-1)$. Recall this is the same as

$$
\begin{aligned}
U_{:,1}^T\left(\frac{1}{n-1}\widetilde{X}\widetilde{X}^T\right)U_{:,1} &= \frac{1}{n-1}U_{:,1}^T\left(USV^TVSU^T\right)U_{:,1}\\
&= \frac{1}{n-1}U_{:,1}^T\left(US^2U^T\right)U_{:,1}\\
&= S_{11}^2/(n-1).
\end{aligned}
$$

4. Let $A \in \mathbb{R}^{m\times n}$. Find maximizers $\vec{x}\in\mathbb{R}^m, \vec{y}\in\mathbb{R}^n$ solving

$$
\begin{array}{ll}
\text{maximize} & \vec{x}^T A\vec{y}\\
\text{subject to} & \|\vec{x}\|_2 = 1,\\
& \|\vec{y}\|_2 = 1.
\end{array}
$$

Also give the maximum value obtained.

Solution. Compute the SVD $A = USV^T$. The maximizers are given by $\vec{x} = U_{:,1}$ and $\vec{y} = V_{:,1}$ with maximum value given by $\sigma_1 = S_{11}$.

To see this is the maximum, note that

$$\vec{x}^T A\vec{y} \leq \|\vec{x}\|\|A\vec{y}\| = \|A\vec{y}\| \leq S_{11},$$

by Cauchy-Schwarz.

5. In the following, assume every day is equally likely to be a birthday, and that there are no leap years.

(a) What is the probability, in terms of n, that at least 2 people in a room of n people have the same birthday?

(b) Give an upper bound, in terms of n and k, that at least k people in a room of n people have the same birthday.

Solution.

(a) $1 - \frac{365\cdot364\cdots(365-n+1)}{365^n}$

(b) The probability that a fixed set of k people all have the same birthday is $\frac{365}{365^k} = 365^{-(k-1)}$. Applying the union bound we obtain an upper bound of $\binom{n}{k}365^{-(k-1)}$. The union bound is the statement that

$$\mathbb{P}\left(\bigcup_n A_n\right) \leq \sum_n \mathbb{P}(A_n),$$

for a finite or infinite sequence of events A_n.

6. If $X \sim \mathcal{N}(0,1)$ then we say that $X^2 \sim \chi_1^2$ (called a chi-squared distribution with 1 degree of freedom). Give the pdf, mean, and variance of the χ_1^2 distribution.

Solution. Let $Y = X^2$. To compute the pdf we use the cdf $F_Y(y)$ of Y for $y \geq 0$:

$$\begin{aligned}
F_Y(y) &= \mathbb{P}(Y \leq y) \\
&= \mathbb{P}(X^2 \leq y) \\
&= \mathbb{P}(-\sqrt{y} \leq X \leq \sqrt{y}) \\
&= \mathbb{P}(-\sqrt{y} < X \leq \sqrt{y}) \\
&= F_X(\sqrt{y}) - F_X(-\sqrt{y}).
\end{aligned}$$

The pdf of Y is given by

$$f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{f_X(\sqrt{y})}{2\sqrt{y}} - \frac{f_X(-\sqrt{y})}{-2\sqrt{y}} = \frac{f_X(\sqrt{y})}{\sqrt{y}} = \frac{e^{-y/2}}{\sqrt{2\pi y}},$$

for $y > 0$ and 0 otherwise.

7. What is the joint pdf of a random vector $X \sim \mathcal{N}(0, I)$ taking values in $\mathbb{R}^n$?

Solution. Since the components are i.i.d. the pdf must factor giving

$$f(x_1, \ldots, x_n) = \prod_{k=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-x_k^2/2} = \frac{e^{-\|x\|_2^2/2}}{(2\pi)^{n/2}},$$

where $x = (x_1, \ldots, x_n)$.

8. Let $A = \begin{bmatrix} 4 & -1 \\ 4 & 1 \end{bmatrix}$ and $\vec{b} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$. Suppose $X \sim \mathcal{N}(0, I)$ takes values in $\mathbb{R}^2$, and let $Y = AX + \vec{b}$.

   (a) What is the distribution of Y?

   (b) What are the marginal distributions of the components of Y?

   (c) Are the components of Y independent?

   (d) What do the contour lines of the joint pdf Y look like?

   Solution.

   (a) $Y \sim \mathcal{N}(\vec{b}, AA^T)$

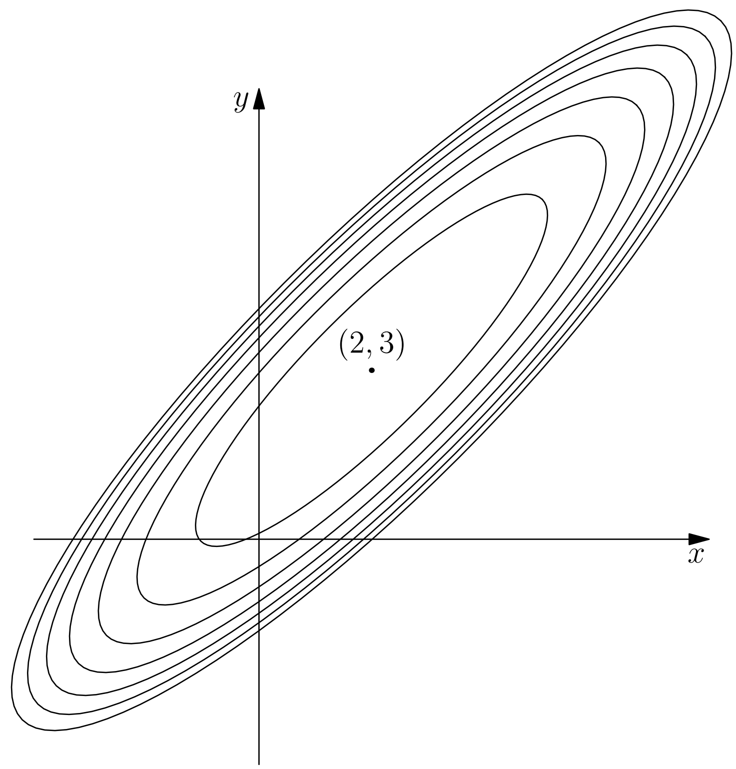
   (b) Since

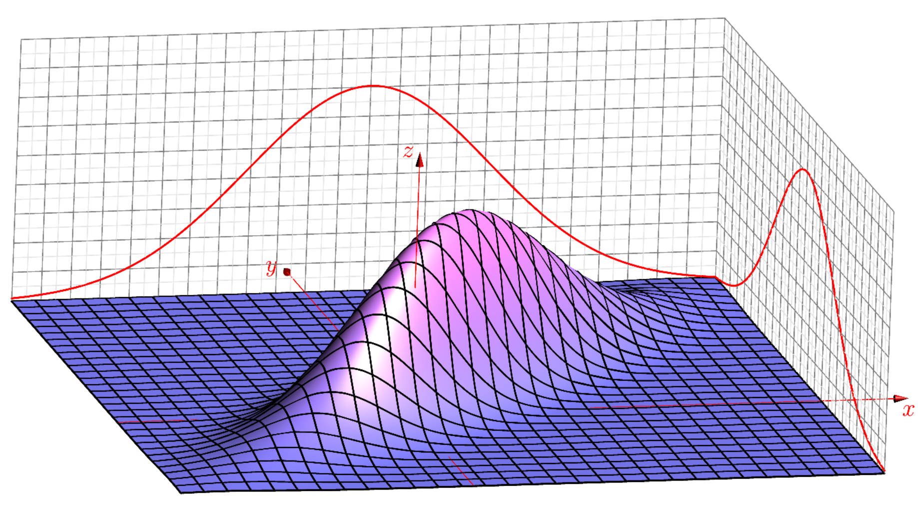
   $$AA^T = \begin{bmatrix} 17 & 15 \\ 15 & 17 \end{bmatrix},$$

   we have $Y[1] \sim \mathcal{N}(2, 17)$ and $Y[2] \sim \mathcal{N}(3, 17)$.

(c) No, as they are positively correlated.

(d) Below we give a contour plot of the joint pdf along with a 3d plot.





This can be understood by computing the SVD of A:

$$A = \begin{bmatrix} 4 & -1 \\ 4 & 1 \end{bmatrix} = \begin{bmatrix} 4 & -1 \\ 4 & 1 \end{bmatrix} \begin{bmatrix} 4\sqrt{2} & 0 \\ 0 & \sqrt{2} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}^T.$$

In general, if $A = USV^T$ then V^T applied to an i.i.d. Gaussian vector fixes the contours, S stretches the contours, and then U rotates the stretched contours. Thus, the resulting contours are always ellipsoids.

9. Let $X_1, \ldots, X_n$ be i.i.d. random variables taking the values $-1, 0, +1$ with probabilities $1/3$ each. Let X denote the random vector in $\mathbb{R}^n$ having X_i as its ith coordinate.

   (a) Compute $E[\|X\|_2^2]$.

   (b) Compute $E[\|X\|_\infty]$.

   (c) Compute the covariance matrix of X.

   Solution.

   (a) $E[\|X\|_2^2] = \sum_{k=1}^n E[X_i^2] = 2n/3$.

   (b) $E[\|X\|_\infty] = 1 - 1/3^n$.

   (c) Let $\Sigma = \mathrm{Cov}(X)$. Then $\Sigma_{ii} = 2/3$ and $\Sigma_{ij} = 0$ for $i \neq j$ by independence.

10. Let X be a random vector taking values in $\mathbb{R}^n$ with mean $\vec{\mu} \in \mathbb{R}^n$ and covariance matrix $\Sigma \in \mathbb{R}^{n \times n}$. If $A \in \mathbb{R}^{m \times n}$ and $\vec{b} \in \mathbb{R}^m$ what are the mean and covariance matrix of $AX + \vec{b}$?

    Solution. The mean is $A\vec{\mu} + \vec{b}$ and the covariance is given by $A\Sigma A^T$.

    To see the mean, note that

    $$E[A_{i,:}X] = E[\sum_{k=1}^n A_{ik}X[k]] = \sum_{k=1}^n A_{ik}E[X[k]] = A_{i,:}E[X],$$

    by the linearity of expectation. Applying this to every row shows $E[AX] = AE[X]$.

    To see the covariance, recall that $\mathrm{Cov}(X) = E[(X - \vec{\mu})(X - \vec{\mu})^T]$. Thus we have

    $$\begin{aligned} \mathrm{Cov}(AX + \vec{b}) &= E[(AX + \vec{b} - (A\vec{\mu} + \vec{b}))(AX + \vec{b} - (A\vec{\mu} + \vec{b}))^T] \\ &= E[(A(X - \vec{\mu}))(A(X - \vec{\mu}))^T] \\ &= E[A(X - \vec{\mu})(X - \vec{\mu})^T A^T] \\ &= AE[(X - \vec{\mu})(X - \vec{\mu})^T]A^T \\ &= A\Sigma A^T, \end{aligned}$$

    by linearity of expectation twice (for A on the left and A^T on the right).