Homework 4 Sample Solutions

Problem 1: Clustering mixture models with PCA

(a) Note that $\mathbb{E}X = 0$, so $\Sigma = \mathbb{E}[XX^\top]$. Moreover, for any $u \in \mathbb{R}^d$, we have

$$\mathbb{E}_{X\sim\mathcal{N}}(u, I)[XX^\top] = \mathbb{E}_{X\sim\mathcal{N}}(0, I)[(X+u)(X+u)^\top] = I + uu^\top \ ,$$

since the cross terms vanish. Thus,

$$\boxed{\Sigma = \mathbb{E}[XX^\top] = \frac{1}{2}\mathbb{E}_{X\sim\mathcal{N}(u,I)}[XX^\top] + \frac{1}{2}\mathbb{E}_{X\sim\mathcal{N}(-u,I)}[XX^\top] = I + uu^\top \ .}$$

Since $\|u\|_2 = 1$, the eigenvalue along u is 2, while every direction orthogonal to u has eigenvalue 1. The first principal component therefore gives $\mathrm{span}(u)$ exactly, with the sign of u undetermined.

(b) Let $\widehat{u}$ be a unit top eigenvector of $\widehat{\Sigma} = \Sigma + E$. For every unit vector z, $|z^\top E z| \le \|E\|_{\mathrm{op}} \le \varepsilon$. Since $\widehat{u}$ maximizes the Rayleigh quotient of $\widehat{\Sigma}$,

$$2 - \varepsilon \le u^\top \widehat{\Sigma} u \le \widehat{u}^\top \widehat{\Sigma} \widehat{u} \le 1 + |\langle \widehat{u}, u\rangle|^2 + \varepsilon.$$

With $\Pi = \widehat{u}\widehat{u}^\top$, rearranging gives

$$\boxed{\|\Pi u\|_2^2 = |\langle \widehat{u}, u\rangle|^2 \ge 1 - 2\varepsilon.}$$

(c) First, project a fresh sample onto $\widehat{u}$. The projected distributions are $N(a, 1)$ and $N(-a, 1)$, where $a = \langle u, \widehat{u}\rangle$. and the sign rule succeeds with probability $\Phi(a)$, where Φ is the standard normal cdf. By (b), $|a|$ is very close to one. Thus, the natural rule is to simple cluster based on the sign of the projected point. However, even if $|a| = 1$, i.e. we had perfect recovery, the components overlap, so some errors are unavoidable.

You were not expected to write this up, but for your own reference, the exact success probability can be expressed as

$$\boxed{\Pr[\text{correct component, up to relabeling}] \ge \Phi\big(\sqrt{\max\{0, 1 - 2\varepsilon\}}\big).}$$

Exact PCA gives success $\Phi(1) \approx 0.8413$. This is also the best possible accuracy even with known means: under the equal priors, the log-likelihood ratio is $2\langle u, X\rangle$, so the optimal rule uses its sign.

Problem 2: Interpreting PCA for genetic data

Start by matching *one-dimensional* projections. For example, A and B put the brown cluster at the same positive horizontal coordinates. B's vertical split into two bands reappears as C's horizontal split. The olive cluster with a long positive tail is on the vertical axis of A and C and the horizontal axis of D. This identifies four distinct score coordinates:

Coordinate	Axes that share it
a	A horizontal, B horizontal
b	A vertical, C vertical, D horizontal
c	B vertical, C horizontal
d	D vertical

Visually, it should be obvious that coordinate d is the largest energy. Let E_a, E_b, E_c, E_d be the marginal energies in the table. The left-column hint gives $E_a + E_b > E_c + E_b$, so $E_a > E_c$. The right-column hint gives $E_a + E_c > E_b + E_d$. Since d is the largest, the right-column hint is possible only if b has the smallest energy, from which we obtain the ordering Thus the order $E_d > E_a > E_c > E_b$, which matches the ordering v_1, v_2, v_3, v_6, respectively. The axes are

Plot	Horizontal	Vertical
A	v_2	v_6
B	v_2	v_3
C	v_3	v_6
D	v_6	v_1

Problem 3: On de-randomizing power method

The key is to start with a nonzero component in the dominant eigendirection, that is, we need $\langle x, u_0 \rangle \neq 0$, where x is the top eigenvector.

(a) Fix an integer $M \geq 2$. Draw the coordinates of g independently and uniformly from $\{1, \ldots, M\}$, and start with $u_0 = g/\|g\|_2$. Every such start is a positive unit vector. Choose j with $\ell_j \neq 0$, and condition on every coordinate of g except g_j. There is at most one value of g_j for which $\ell^\top g = 0$. Therefore

$$\boxed{\Pr[Pu_0 \neq 0] \geq 1 - 1/M.}$$

We may make this probability as high as desired by increasing M. If the dominant eigenvalue is negative, the iterates can alternate in sign, but their absolute alignment with x still tends to 1.

(b) Start deterministically at $\boxed{u_0 = \mathbf{1}/\sqrt{d}}$, where $\mathbf{1}$ is the all ones vector. To see why it works, take any successful nonnegative start u from (a). Since u is a unit vector, we have $0 \leq u \leq \mathbf{1}$ coordinatewise. Nonnegativity of A implies

$$0 \leq A^t u \leq A^t \mathbf{1} \qquad \text{coordinatewise for every } t.$$

Take Euclidean norms, divide by $|\lambda_1|^t$, and pass to the limit. This gives $0 < \|Pu\|_2 \leq \|P\mathbf{1}\|_2$. Hence the all-ones start has a nonzero dominant component, so power iteration converges to the top eigendirection.