Lecture 6

Principal component analysis

Suppose we have a matrix $\widetilde{X} \in \mathbb{C}^{n \times p}$ whose rows are centered observations $x_1 - \overline{x}, \ldots, x_n - \overline{x}$ and whose columns correspond to features:[1]

$$\widetilde{X} = \begin{pmatrix} \left(x_1 - \overline{x}\right)^\top \\ \left(x_2 - \overline{x}\right)^\top \\ \vdots \\ \left(x_n - \overline{x}\right)^\top \end{pmatrix} \tag{6.1}$$

If $v \in \mathbb{C}^p$ is a unit vector, the matrix $\widetilde{X}\overline{v}$ is a list of scalar projections of $\widetilde{X}$'s rows onto v.

$$\widetilde{X}\overline{v} = \begin{pmatrix} \left(x_1 - \overline{x}\right)^\top \overline{v} \\ \left(x_2 - \overline{x}\right)^\top \overline{v} \\ \vdots \\ \left(x_n - \overline{x}\right)^\top \overline{v} \end{pmatrix} = \begin{pmatrix} \left\langle \left(x_1 - \overline{x}\right), v \right\rangle \\ \left\langle \left(x_2 - \overline{x}\right), v \right\rangle \\ \vdots \\ \left\langle \left(x_n - \overline{x}\right), v \right\rangle \end{pmatrix} \tag{6.2}$$

Each inner product can be interpreted as having a factor of $\cos\theta$, where θ is the angle between the two vectors. Then $\widetilde{X}\overline{v}$ is larger when θ tends to be small, viz. when v points in the prevailing direction of deviation from $\overline{x}$. We can find v_1 capturing the direction of greatest variation, v_2 an orthogonal direction with second-greatest variation, etc. using the SVD. Define the covariance matrix $Q \in \mathbb{C}^{p \times p}$ as follows:

$$Q = \frac{1}{m-1} \left(\widetilde{X}\right)^* \widetilde{X} \tag{6.3}$$

This can be seen as the sample average value of xx^*.[2]

- Diagonal entry Q_{jj}[3] is the average squared distance that feature p lands from its average value. This called the variance of feature j.
- Off-diagonal entry Q_{jk} is what you expect on the average when you multiply feature j by the complex conjugate of feature k. It captures the correlation between feature j and feature k. If $Q_{jk} \neq 0$, then features j and k tend to move together.
  - If Q_k is positive, that means that features j and k tend to deviate in the same direction. When the former goes up, the latter goes up.

[1] Studying PCA in the context of complex observation would be considered exotic in the practice of statistics, but real-world PCA works just as well if you skip all the conjugation.

[2] Except the denominator is $m-1$ instead of m for statistical reasons. This is called Bessel's correction.

[3] Q_{-1}?

- If Q_k is negative, then features j and k tend to deviate in opposite directions. When the former goes up, the latter goes up.
- If Q_k is pure imaginary, then features j and k always move together, but at some angle in the complex plane.[4] When the former goes east, the latter goes north or south.

Let λ_i be the ith greatest eigenvalue of Q, and v_i a unit eigenvector that satisfies $Qv_i = \lambda_i v_i$.

- v_i is the ith **principal component** of $\widetilde{X}$. Projection onto v_1 gets more scalar variance out of $\widetilde{X}$ than any other direction. If a scatter plot of the rows of $\widetilde{X}$ looks like an ellipse in two dimensions, then v_1 is the semi-major axis and v_2 is the semi-minor axis.
- λ_i is the variance of $\widetilde{X}$ after projection onto the direction v_i. If a scatter plot of the rows of $\widetilde{X}$ looks like an ellipse in two dimensions, then $\sqrt{\lambda_1}$ is proportional to the length of the semi-major axis, and $\sqrt{\lambda_2}$ is proportional to the length of the semi-minor axis.

Scatter plots of empirical data come in all shapes and sizes, but after you project $\widetilde{X}$ onto its leading principal components (by taking inner products e.g. ${x_i}^\top \overline{v_j}$), the scatter plots all look quite the same, at least in these ways:

- The points are centered at the origin.
- The point cloud is longest along the axis of the foremost principal component.

6.1 Example: measuring an impedance by hand

Suppose that you are abducted by aliens. You are presented with an unknown linear circuit component, an AC voltage source, an oscilloscope, and a graphing calculator capable of linear algebra. The aliens will set you free, but only if you can tell them what the impedance Z is at a frequency ω.

You beg them to let you into your lab at Berkeley where you have a instrument that measures impedance, but they say no. Here is how you might get a pretty good guess:

1. Wire up your oscilloscope to plot voltage and current while you use your AC voltage source to induce a voltage $v(t) = V\cos(\omega t)$ over the mystery component.
2. Write down the current waveform as $i(t) = I\cos\big(\omega t + \phi\big)$.
3. Now you have a pair of phasors $x = (\tilde{V}, \tilde{I})$, where $\tilde{V} = V$ and $\tilde{I} = Ie^{j\phi}$.
4. Instead of immediately reporting the ratio $Z = \tilde{V}/\tilde{I}$, collect more data at a range of voltages to be extra sure.
5. Now you have a long list of points $\{x_1, x_2, \ldots, x_n\}$ where $x_i = (\tilde{V}_i, \tilde{I}_i)$.
6. Consult the aliens, who are able to visualize data in 4 spatial dimensions, and confirm that your data points do indeed look like a line in $\mathbb{C}^2$.
7. Perform PCA on your data to find the first principal component (a, b). When $\tilde{V}$ moves in a direction of a, $\tilde{I}$ moves in the direction b.
8. Report the slope a/b as your impedance estimate.

[4] The interpretation of this case, I admit, is quite bizarre. I believe it is never part of the statistical form of PCA, which works with real data only.