Crib 5

05 PCA, CCA

by Alvin Wan . alvinwan.com/cs189/fa17

Note that in the objective functions below, you may choose to featurize your data i.e., replace all x_i with $\phi(x_i)$

1 Principal Components Analysis (PCA)

1. Objective: minimizing projection error $\min_u \|y - Xu\|_2^2$ or maximize Rayleigh quotient $\max_{u:\|u\|=1} \frac{u^T XX^T u}{u^T u}$ *(see proof of equivalence in the appendix)*
2. Motivation: Data is largely explained by its components in a lower-dimensional representation.
3. Probabilistic Interpretation: Fitting a Gaussian model to our data, where $x_i = a_i b_i + z_i$ for Gaussian noise z_i. a_i are fixed and our goal is to find b_i.

2 Cross Correlation Analysis (CCA)

See my longer note on CCA for more detailed motivation. See the course notes for more breadth.

1. Objective: $\max_{u,v\in\mathbb{R}^d} \mathrm{cor}(u^T x, v^T y)$, where $\mathrm{cor}(a, b) = \frac{\mathrm{cov}(a,b)}{\sqrt{\mathrm{var}(a)\mathrm{var}(b)}}$
2. Motivation: The directions of most variance in our data do not contribute to correlation between x and y.
3. Probabilistic interpretation: Take a matrix that whitens X, Y, then a matrix that decorrelates the two.

3 Proof: Minimizing Projection Error is Maximizing Rayleigh Quotient

Taken from my previous notes.

This is reminiscent of regression: when we run linear regression, the error runs along the y-axis. For PCA, the error is the minimum distance from the point to the line, making the error perpendicular to our line. In other words, if our decision boundary is defined by

$$\{w^T z = 0\} = L(w)$$

We have that distance is defined is to be the following. Note that the projection of x_i onto w is $\text{proj}_w x_i = \frac{\langle w, x_i \rangle}{\langle w, w \rangle} = \frac{w^T x_i}{\|w\|^2} w$. To take the distance to our line, we then consider the magnitude of the difference between the projection with x_i.

$$\textsc{Dist}(x_i, L(w) = \|x_i - \frac{w^T x_i}{\|w\|^2} w\|^2$$

As a result, our objective is simply the sum of all these distances to the line. In the first step, we use $(a-b)^2 = a^2 - 2ab + b^2$. In the second, we use the fact that $w^T x_i$ is a scalar and that $x_i^T w = (w^T x)^T$ is the same scalar. In the fourth, we know that $\frac{w}{\|w\|}$ is a unit vector and that $\frac{w^T x_i}{\|w\|}$ is a scalar.

$$\begin{aligned}
&\textsc{Minimize}_w \sum_{i=1}^n \|x_i - \frac{w^T x_i}{\|w\|^2} w\|^2 \\
&= \sum_{i=1}^n \|x_i\|^2 - 2x_i^T \frac{w^T x_i}{\|w\|^2} w + \|\frac{w^T x_i}{\|w\|^2} w\|^2 \\
&= \sum_{i=1}^n \|x_i\|^2 - 2\frac{(w^T x_i)^2}{\|w\|^2} + \|\frac{w^T x_i}{\|w\|^2} w\|^2 \\
&= \sum_{i=1}^n \|x_i\|^2 - 2(\frac{w^T x_i}{\|w\|})^2 + \|\frac{w^T x_i}{\|w\|} \frac{w}{\|w\|}\|^2 \\
&= \sum_{i=1}^n \|x_i\|^2 - 2(\frac{w^T x_i}{\|w\|})^2 + (\frac{w^T x_i}{\|w\|})^2 \\
&= \sum_{i=1}^n \|x_i\|^2 - (\frac{w^T x_i}{\|w\|})^2
\end{aligned}$$

Since x_i are fixed, then $\|x_i\|_2^2$ are fixed. We can reformulate this as maximization problem, considering only the second term.

$$\begin{aligned}
&\textsc{Maximize}_w \sum_i (\frac{w^T x_i}{\|w\|})^2 \\
&= \frac{\sum_i w^T x_i x_i^T w}{\|w\|^2} \\
&= \frac{w^T X X^T w}{\|w\|^2} \\
&= (\frac{w}{\|w\|})^T X X^T \frac{w}{\|w\|}
\end{aligned}$$

$\frac{w}{\|w\|}$ are unit vectors, so we can consider $u \in \mathbb{R}^d$, where $\|u\|_2 = 1$.

$$\textsc{Maximize}_{u:\|u\|_2=1} u^T X X^T u$$

Note this objective function is precisely the formulation for PCA.