HW 9

Evan Ott
UT EID: eao466

November 14, 2016

Notes from class 11/14

PCA

Idea from PCA is to summarize most of the information in a smaller number of covariates. To do this, you maximize the variance along the principal axis. This turns out to be basically just an eigenvalue problem, with the most important axis being the one with the largest eigenvalue, etc. Because of this, scaling of variables is critical. What's even nicer is that we can find these from:

$$\begin{aligned}\text{cov} &= X^\top X/(n-1) = VLV^\top \\ X &= USV^\top \\ \text{cov} &= V\frac{S^2}{n-1}V^\top\end{aligned}$$

For more on PCA, see James et al. (2013)

PMD

In PCA, V is not guaranteed at all to be sparse. That's where the new paper comes in, see Witten et al. (2009). Now, we'll construct r factors by

$$\underset{\hat{X}\in M(r)}{\text{argmin}} \left\| X - \hat{X} \right\|_F^2 = \sum_{k=1}^{r} d_k u_k v_k^\top$$

where $\|\cdots\|_F^2$ is the Frobenius norm, the sum of the squares of all matrix elements.

Now, we will set out to solve

$$\begin{aligned}\underset{d,u,v}{\text{minimize}}\ & \frac{1}{2}\left\|X - duv^\top\right\|_F^2 \\ \text{subject to}\ & \|u\|_2^2 = 1 \quad \|v\|_2^2 = 1 \\ & \|u\|_1 \le c_1 \quad \|v\|_1 \le c_2\end{aligned}$$

The trick is that we end up being able to solve for d trivially so that we really only alternate between solving for u and v until convergence.

A small note: if we dropped the $\mathcal{L}_1$ constraint, then update solution for u would be $\frac{Xv}{\|Xv\|_2}$ because we just want to maximize $u^\top Xv$ (go in same direction, see Theorem 2.1 in the paper), but have unit length (scale by $\mathcal{L}_2$ norm).

When we then include the $\mathcal{L}_1$ constraint, that's where we get the soft-thresholding piece.

Another small note: generating a rank-k approximation will not be guaranteed to reproduce the original rank-k X matrix because this approximation is *sparse*.

References

Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An introduction to statistical learning*, volume 6. Springer, 2013. URL http://www-bcf.usc.edu/~gareth/ISL/.

Daniela M Witten, Robert Tibshirani, and Trevor Hastie. A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics*, 10:515–534, 2009. URL https://faculty.washington.edu/dwitten/Papers/pmd.pdf.