Class 10 Review Notes

AI & Machine Learning

Fall 2023 - Laurie Ye

10 Principle Components for Images

10.1 Introduction

Even fairly small images contain a large number of pixels. For example, for greyscale images:

- 200×200 image = 40,000 pixels
- 400×400 image = 160,000 pixels
- 500x500 image = 250,000 pixels

Since, in an image, the pixels are the "features," this means that images have a huge number of x-variables. As a result, when working with images, it is often the case that you have far fewer observations than you have x-variables.

10.2 Problem

- Suppose I have a 400×400 image of each student in my class. This will mean that I would generally have 50 or fewer images. But each image has 160,000 pixels. So the sample size of 50 is vastly smaller than the number of x-variables (160,000).
- In fact, the number of pixels is so large that it would be very difficult to obtain a sample of images large enough to exceed the number of x-variables.
- In addition, a large number of x-variables can make problems too big to fit into computer memory and can make computations slow. For example, for a 400×400 image, it may be difficult to load the variance-covariance matrix of the pixels into memory (160,000 × 160,000 = 25,600,000,000 = 25.6 billion entries).
- So, in this way images are quite special. They are also special in that all of the variables are measure the same thing and are on the same scale.

10.3 Trick

Setup

- Let X be an $n \times p$ matrix representing a sample of n images with p pixels.
- Each row of X contains the data for one image. Let X_c be the "centered" X; that is subtract the column average from each column of X.

- Then $\frac{X_c'X_c}{(n-1)} = \hat{\Sigma}$, the $p \times p$ (huge) sample variance-covariance matrix of the pixels.

The Trick

$X_c'X_c$ is huge ($p \times p$). X_cX_c' is small ($n \times n$). Can we relate the eigenvectors and eigenvalues of $\frac{X_c'X_c}{(n-1)} = \hat{\Sigma}$ to those of the small matrix X_cX_c'?

The answer is that WE CAN.

Suppose that λ and v are an eigenvalue, eigenvector pair of the matrix X_cX_c'.

That is $X_cX_c'v = \lambda v$.

Then

$$\frac{X_c'X_c}{n-1}X_c'v = \frac{\lambda}{n-1}X_c'v$$
$$\frac{X_c'X_c}{n-1}X_c'v = \frac{\lambda}{n-1}X_c'v$$
$$\hat{\Sigma}X_c'v = \frac{\lambda}{n-1}X_c'v$$

Now $\hat{\Sigma}X_c'v = \frac{\lambda}{n-1}X_c'v$ means that $X_c'v$ is an eigenvector of $\hat{\Sigma}$. Well, almost. We should worry about the convention that eigenvectors have length 1 and are orthogonal.

For length, divide by the norm

$$\|X_c'v\| = \sqrt{v'X_cX_c'v}$$

So,

$$\hat{\Sigma}\frac{X_c'v}{\|X_c'v\|} = \frac{\lambda}{n-1}\frac{X_c'v}{\|X_c'v\|}$$

Now check orthogonality: Suppose v_1 and v_2 are two eigenvectors of X_cX_c'. So, $v_1'v_2 = 0$.

We need

$$\left(\frac{X_c'v_1}{\|X_c'v_1\|}\right)'\left(\frac{X_c'v_2}{\|X_c'v_2\|}\right) = 0$$

$$\left(\frac{X_c'v_1}{\|X_c'v_1\|}\right)'\left(\frac{X_c'v_2}{\|X_c'v_2\|}\right) = 0$$

implies

$$\left(\frac{X_c'v_1}{\|X_c'v_1\|}\right)'\left(\frac{X_c'v_2}{\|X_c'v_2\|}\right) = 0$$

or

$$v_1'X_cX_c'v_2 = 0$$

But $X_c X_c' v_2 = \lambda_2 v_2$. So,

$$v_1' X_c X_c' v_2 = v_1' \lambda_2 v_2 = \lambda_2 v_1' v_2$$

But, $v_1' v_2 = 0$. So the eigenvectors $\frac{X_c' v}{\|X_c' v\|}$ are orthogonal.

10.4 Summary

Suppose that λ and v are an eigenvalue, eigenvector pair of the matrix $X_c X_c'$.

Then, $\frac{\lambda}{n-1}$ and $\frac{X_c' v}{\|X_c' v\|}$ are an eigenvalue, eigenvector pair of $\hat{\Sigma} = \frac{X_c' X_c}{(n-1)}$.

Note: there will be n such pairs. All other eigenvalues of $\hat{\Sigma}$ will be 0.

10.5 Comment

Note that while

$$\text{Var}(X) = \hat{\Sigma} = \frac{X_c' X_c}{(n-1)}$$

it is NOT the case that

$$\text{Var}(X') = \hat{\Sigma}_{x'} = \frac{X_c X_c'}{(p-1)}$$

because the columns are not centered correctly. So don't try to compute $X_c X_c'$ by using $\text{Var}(X')$ or $\text{Var}(X_c')$.

10.6 The Punch Line

- If I am in the situation I described earlier where I have about 50 images of size 400×400 pixels, instead of having to do an eigenvalue-eigenvector decomposition of a $160,000 \times 160,000$ variance-covariance matrix, I can do an eigenvalue-eigenvector decomposition of a 50×50 matrix instead.

10.7 Reference

Goizueta Business School-Emory University: Professor George S. Easton