Definition of PCA

Yifeng Tao

April 17, 2016

There are two ways to deduce PCA formulation. Namely, the way based on maximum variance or minimum error definition.

I refered to Bishop's book and found it has **logic errors**:

- The proof using induction assumes that when we try to extend the dimension from M to M+1, the M vectors remains unchanged.
- In addition, to achieve the minimum error goal, we do **NOT** have to choose the top M eigenvectors of feature matrix. The vectors we choose only need to satisfy: They form the same subspace of top M eigenvectors.

Here we begin our own proof:

Assume we choose M direction: $u_1, ..., u_M$ in the D-dimension space, where $\{u_m\}_{m \in [M]}$ are orthogonal. That is: $u_i^T u_j = \delta_{ij}$.

Assume $U = [u_1, ..., u_M]$, and

$$S = \frac{1}{N}\sum_{n=1}^{N}(x_n - \bar{x})(x_n - \bar{x})^T.$$

1 Maximum Variance Formulation

We want to maximize variance:

$$J = \frac{1}{N}\sum_{n=1}^{N}||U^T x_n - U^T \bar{x}||^2 = \frac{1}{N}\sum_{n=1}^{N}\sum_{m=1}^{M}||u_m^T x_n - u_m^T \bar{x}||^2 = \sum_{m=1}^{M}\frac{1}{N}\sum_{n=1}^{N}||u_m^T x_n - u_m^T \bar{x}||^2 = \sum_{m=1}^{M} u_m^T S u_m.$$

Our problem becomes:

$$\max_{u_m, m \in M} J = \sum_{m=1}^{M} u_m^T S u_m,$$

s.t.

$$u_i^T u_j = \delta_{ij}.$$

The Lagrange multiplier of this optimization problem is:

$$L = \sum_{m=1}^{M} u_m^T S u_m - \sum_{1 \le i \le j \le M} t_{ij}(u_i^T u_j - \delta_{ij}).$$

Partial derivative to u_m should equal to 0:

$$\frac{\partial L}{\partial u_m} = 2Su_m - \sum_{i=1}^{M} k_{mi}u_i = 0.$$

We left multiply u_m^T to it. Note that $u_m^T u_i = \delta_{mi}$, we get:

$$u_m^T S u_m = k_{mm}/2 = \lambda_m.$$

Then we left multiply u_m to both side:

$$Su_m = \lambda_m u_m.$$

Thus, u_m should be the eigenvectors of S, λ_m is the corresponding eigenvalues.

Our problem becomes:

$$\max_{u_m, m\in M} J = \sum_{m=1}^{M} u_m^T S u_m = \sum_{m=1}^{M} \lambda_m,$$

s.t.

$$u_i^T u_j = \delta_{ij}, u_m \text{ are eigenvalues of } S.$$

From all the eigenvetors of S, we choose the top M eigenvectors. $Q.E.D.$

2 Minimum Error Formulation

We can have a complete orthogonal set of D-dimension basis vectors $\{u_i\}_{i\in[D]}$.

Each point can be represented as:

$$x_n = \sum_{i=1}^{D} \alpha_{ni} u_i$$

And multiply u_i^T on the left, we have:

$$x_n = \sum_{i=1}^{n} (x_n^T u_i) u_i$$

Now, we only select M of these u_i to represent all the points. We approximate them into:

$$\tilde{x_n} = \sum_{i=1}^{M} z_{ni} u_i + \sum_{i=M+1}^{D} b_i u_i$$

We want to minimize

$$J = \frac{1}{N} \sum_{n=1}^{N} ||x_n - \tilde{x_n}||^2$$

First minimize with respect to z_{ni}, use Lagrangian function, we get:

$$z_{ni} = x_n^T u_i$$

We can also minimize with respect to b_i, use Lagrangian function, we get:

$$b_i = \bar{x}^T u_i$$

Thus,

$$x_n - \tilde{x_n} = \sum_{i=M+1}^{D} ((x_n - \bar{x})^T u_i) u_i$$

So, our opimization problem become minimizing

$$J = \frac{1}{N} \sum_{n=1}^{N} \sum_{i=M+1}^{D} (x_n^T u_i - \bar{x}^T u_i)^2 = \sum_{i=M+1}^{D} u_i^T S u_i$$

The explanation is similar to former (in section 1), we finally get $\{u_i\}_{i=M+1,\ldots,D}$ should be the least eigenvectors of S. Thus, if we choose top M eigenvectors of S, it will satisfy all. $Q.E.D.$

However, we have to note that, $\{u_i\}_{i\in[M]}$ don't have to be the top M eigenvectors, they only have to satisfy that: They form the same subspace of top eigenvectors.

Acknowlegement

Thanks a bunch to Libin Liang for his suggestion and reading on the original proof.

References

[1] Bishop et al. Pattern Recognition and Machine Learning. Springer Science+Business. 2006