PCA Vectorized Derivation

Bikramjot Hanzra

In PCA, we want to find a set of orthogonal basis vectors that maximize variance of the projections of the data along each subsequent dimension. Let $\mathbf{X}$ be the $n \times d$ matrix where each row is a data point and $\mathbf{w}$ be the $d \times 1$ dimension column vector. The optimization problem can be formulated as –

$$J(\mathbf{w}) = \arg\max_{\mathbf{w}} ||\mathbf{X} \cdot \mathbf{w}||^2 \text{subjected to} ||\mathbf{w}|| = 1 \tag{1}$$

Introducing lagrange multiplier λ we get –

$$J(\mathbf{w}) = \arg\max_{\mathbf{w}} ||\mathbf{X} \cdot \mathbf{w}||^2 + \lambda(1 - ||\mathbf{w}||^2) \tag{2}$$

$$J(\mathbf{w}) = \arg\max_{\mathbf{w}} (\mathbf{X} \cdot \mathbf{w})^\intercal \cdot (\mathbf{X} \cdot \mathbf{w}) + \lambda(1 - \mathbf{w}^\intercal \cdot \mathbf{w}) \tag{3}$$

$$J(\mathbf{w}) = \arg\max_{\mathbf{w}} \mathbf{w}^\intercal \cdot \mathbf{X}^\intercal \cdot \mathbf{X} \cdot \mathbf{w} + \lambda(1 - \mathbf{w}^\intercal \cdot \mathbf{w}) \tag{4}$$

The Jacobian $\nabla J(\mathbf{w})$ is given by –

$$\nabla J(\mathbf{w}) = (\mathbf{X}^\intercal \mathbf{X} \cdot \mathbf{w})^\intercal + \mathbf{w}^\intercal \cdot \mathbf{X}^\intercal \cdot \mathbf{X} - \lambda \mathbf{w}^\intercal - \lambda \mathbf{w}^\intercal \tag{5}$$

$$\nabla J(\mathbf{w}) = 2(\mathbf{X}^\intercal \mathbf{X} \cdot \mathbf{w})^\intercal - 2\lambda \mathbf{w}^\intercal \tag{6}$$

The minima will occur where $\nabla J(\mathbf{w}) = 0$.

$$2(\mathbf{X}^\intercal \mathbf{X} \cdot \mathbf{w})^\intercal - 2\lambda \mathbf{w}^\intercal = 0 \tag{7}$$

$$(\mathbf{X}^\intercal \mathbf{X} \cdot \mathbf{w})^\intercal = \lambda \mathbf{w}^\intercal \tag{8}$$

$$\mathbf{X}^\intercal \mathbf{X} \cdot \mathbf{w} = \lambda \mathbf{w} \tag{9}$$

$$\tag{10}$$

From Equation 9 it is easy to see that the orthogonal basis vectors are the eigenvalues of the scatter matrix ($\mathbf{X}^\intercal \mathbf{X}$). This can be implemented in MATLAB using –

```
[eigvec, eigval] = eig(X'*X)
```