Notes on Dimensionality Reduction

[W/ content from MML Book (Deisenroth, Faisal, Ong)]

Given: $\mathcal{X} = \{x_1, x_2, \ldots, x_N\}$

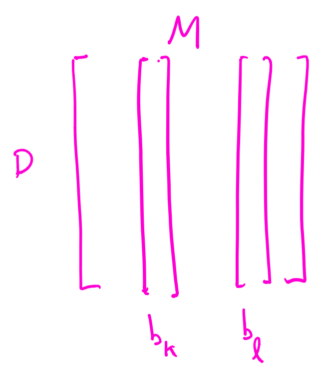
$x_i \in \mathbb{R}^D \qquad \bar{X} = \varnothing$

Key Assumption: There exists a low-dimensional representation

$$z_i = B^T x_i \in \mathbb{R}^M \qquad B \in \mathbb{R}^{D\times M}, \quad M \ll D$$

PCA assumes a linear mapping from $X \to z_i$

This is a Projection and B is an orthonormal matrix.



$\forall \ell, k \quad \ell \neq k$

$b_\ell^T \cdot b_k = \varnothing \longleftrightarrow$ orthogonal

$\forall k$

$b_k^T \cdot b_k = 1$

"bottleneck"

Encode $z_i = B^T x_i$ from $D \to M$ dims

Decode $\tilde{x}_i = \underbrace{B}_{\langle D\times M\rangle}\underbrace{z_i}_{\langle M\times 1\rangle} = B(B^T x_i) \quad M \to D$

So z_i is a compressed representation of x_i and $\tilde{x}_i$ is a reconstruction.

What should our objective function be here?

One perspective: Maximize variance in the lower dimensional space $\tilde{x}$.

↪ Leads to cool connection with Eigenvalues!

Consider first dimension, $\tilde{x}_1$.

$$V_1 = \frac{1}{N} \sum_i^N (\underbrace{b_1^T}_{1 \times D} \underbrace{x_i}_{D \times 1})^2$$

* Remember: Assuming $\bar{x} = \emptyset$

$$= b_1^T \underbrace{\left(\frac{1}{N} \sum_i x_i x_i^T \right)}_{S\text{: covariance matrix}^*} b_1$$

$$= b_1^T S b_1$$

Want to find b_1 to maximize this, but this is trivial without additional constraint

$$\boxed{\max_{b_1} b_1^T S b_1 \quad \text{s.t. } \|b_1\|^2 = 1}$$

We have a constrained problem $\longrightarrow$ Lagrangian Opt.!

Re-write as

$$L(b_1, \lambda) = b_1^T S b_1 + \lambda(1 - b_1^T b_1)$$

$$\frac{\partial L}{\partial b_1} = 2b_1^T S - 2\lambda b_1^T \qquad \frac{\partial L}{\partial \lambda} = 1 - b_1^T b_1$$

Set both to $\emptyset$.

$$\left.\begin{array}{l} \longrightarrow S b_1 = \lambda b_1 \\ \quad b_1^T b_1 = 1 \end{array}\right\}$$

By definition b_1 is an eigenvector of S (λ is the eigenvalue)!

In general, λ is the first principal component.

$$\tilde{x}_i = b_1 z_{1i} = b_1(b_1^T x_i) \in \mathbb{R}^D$$

D dimensions   1 dimension!

To "compress" the information not captured by b_1, we can repeat this procedure on

$$X - \underbrace{b_1 b_1^T X}_{\text{reconstruction}}$$

This will yield b_2, λ_2 where b_2 maximizes remaining variance.

This iterative approach actually not necessary; the Eigenvectors of S are all we need!

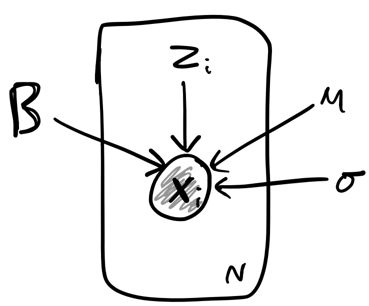
$$B = \begin{bmatrix} b_1 & b_2 & \cdots & b_M \end{bmatrix}$$

Rank in order of λ_k

Take top M.

Let's implement This! [See In-Class exercise]

A probabilistic perspective on PCA



$$z_i \sim N(0, I)$$

$$x_i \mid z_i \sim N(Bz_i + \mu, \sigma^2 I)$$

$$P(x, z \mid B, \mu, \sigma^2)$$

$$= P(x \mid z, B, \mu, \sigma^2)\, P(z)$$

We can use this to generate images from the latent space!

[See notebook]

Projection perspective

$$\tilde{x}_i = B z_i \qquad B \text{ again orthonormal}$$

Want to minimize reconstruction error

$$J_M \overset{def}{=} \frac{1}{N} \sum_{i=1}^{N} \|x_i - \tilde{x}_i\|^2$$

What are the optimal coordinates z_i for x_i w.r.t. B?

$$\frac{\partial J_M}{\partial z_{ji}} = \frac{\partial J_M}{\partial \tilde{x}_i} \frac{\partial \tilde{x}_i}{\partial z_{ji}}$$

$$\frac{\partial J_M}{\partial \tilde{x}_i} = -\frac{2}{N}(x - \tilde{x}_i)^T \in \mathbb{R}^{1 \times D}$$

$$\frac{\partial \tilde{x}_i}{\partial z_{ji}} = \frac{\partial}{\partial z_{ji}} \left(\sum_{m=1}^{M} z_{mi} b_m \right) = b_j$$

So

$$\frac{\partial J_M}{\partial z_{ji}} = -\frac{2}{N}(x_i - \tilde{x}_i)^T b_j$$

$$= -\frac{2}{N}\left(x_i - \sum_{m=1}^{M} z_{mi} b_m\right)^T b_j \qquad (b_m^T \cdot b_j = \emptyset \text{ if } m \neq j)$$

$$= -\frac{2}{N}\left(x_i^T b_j - z_{ji}\, b_j^T b_j\right) = -\frac{2}{N}\left(x_i^T b_j - z_{ji}\right)$$

Set to $\emptyset \rightarrow z_{ji} = b_j^T x_i$

A similar argument for choice of B (basis) can be made, yielding again the M largest Eigenvectors (see reading!)