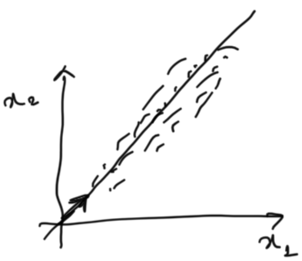
# Principal Component Analysis (PCA)

**Setup**: Given $\mathcal{D} := \{x_1, \ldots, x_n\}$, $x_i \in \mathbb{R}^d$, $d >> 1$

**Goal**: Encode/compress the data in a low-dimensional representa




$$\underset{\in \mathbb{R}^q}{z_i} = f(\underset{\in \mathbb{R}^d}{x_i}), \quad z_i \in \mathbb{R}^q, \quad q << d$$

$$\xrightarrow{\text{PCA}} \quad z_i = \underset{?}{W}^T x_i$$

- Hotelling transfor
- Karhunen-Loeve de
- Proper orthogonal de
- SVD / PCA

- **Maximum variance formulation of PCA (Hotelling 1933):**

Consider a 1-dimensional sub-space, i.e. $\boxed{q = 1}$

Define $u_1 \in \mathbb{R}^d$ to be a coordinate of a 1d subspace

with $\|u_1\|_2 = u_1^T u_1 = 1$

Each data point $x_i \in \mathbb{R}^d$ can be projected on the sub-space spanned by $u_1$ as: $\underset{1 \times d}{u_1^T} \underset{d \times 1}{x_i}$ (scalar)

Mean of all projected data: $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, $u_1^T \bar{x}$

Variance: $\frac{1}{n}\sum_{i=1}^{n}\{u_1^T x_i - u_1^T \bar{x}\}^2 = u_1^T S u_1$

$$\boxed{\underset{d \times d}{S} := \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})^T}$$ sample covariance

We seek the coordinate $u_1$ that captures most variance in

We seek the

the observed data :

$$u_1^* = \arg\max_{u_1} \left\{ \underbrace{u_1^T S u_1} + \lambda (1 - u_1^T u_1) \right\} := \mathcal{L}(u_1$$

Take derivative and set to zero :

$$\nabla_{u_1} \mathcal{L}(u_1) = 0 \implies S u_1 = \lambda u_1$$

$$\implies u_1^T S u_1 = \lambda_1$$

<u>Conclusion</u> : The optimal $u_1^*$ corresponds to the eigenvector of $S$, corresponding to the maximum eigenvalue of $S$.

What about $u_2^*$ ?

$$\begin{cases} u_2^* = \arg\max_{u_2} \left\{ u_2^T S u_2 + \lambda_2 (1 - u_2^T u_2) \right\} \\ \text{s.t. } u_1^* \perp u_2 \end{cases}$$

$\implies$ $u_2^*$ corresponds to the eigenvector of $S$ with the second largest eigenvalue.

<u>Practical implementation</u> :

Given data $X \in \mathbb{R}^{n \times d}$

i.) Normalize : $\hat{X} = X - \mathbb{E}[X]$ , $\mathbb{E}[\hat{X}] = 0$

ii) $\underset{d\times d}{S} = \frac{1}{n}\hat{X}^T\hat{X}$

iii) Compute the SVD of $S = \underset{d\times d}{W}\,\underset{d\times d}{\Lambda}\,\underset{d\times d}{W^T}$

iv) Sort the eigenvalues and eigenvectors in decreasing order; and keep the first $q \ll d$ of them.

Encoding : $\boxed{\underset{n\times q}{Z} = \underset{n\times d}{X}\,\underset{d\times q}{W}}$ , $q \ll d$

Decoding : $\boxed{\underset{n\times d}{X} = \underset{n\times q}{Z}\,\underset{q\times d}{W^T}}$ ←

---

Probabilistic PCA :

Setup : Given $D := \{x_1, ---, x_m\}$, $x_i \in \mathbb{R}^d$

Assumption : There exist a set of latent variables $Z \in \mathbb{R}^q$, $q<$ that effectively summarize the data $X \in \mathbb{R}^{n\times d}$.

Prior : $p(z) \sim \mathcal{N}(0, I)$ ✓.

Likelihood : $p(x|z) = \mathcal{N}(x \mid zW^T + \mu, \sigma^2 I)$ ✓

Parameters : $\theta := \{W, \mu, \sigma^2\}$

$$p(x) = \int p(x,z)\,dz = \int \overbrace{p(x|z)}^{\mathcal{N}}\,\overbrace{p(z)}^{\mathcal{N}}\,dz$$

Linear

$$\Rightarrow \boxed{p(x) = \mathcal{N}(x \mid \mu, G)}, \text{ where } G = WW^T + \sigma^2 I$$

Since :

$$\mathbb{E}[x] = \mathbb{E}[\overset{0}{\cancel{z}} W^T + \mu + \overset{0}{\cancel{\epsilon}}] = \mu$$

$$\text{Cov}[X] = \dots = WW^T + \sigma^2 I$$

Posterior distribution :

$$p(z \mid x) = \frac{p(x \mid z)\, p(z)}{p(x)} = \mathcal{N}(z \mid m^{-1}(X - \mu) W, \sigma^2 m$$

where $M := W^T W + \sigma^2 I$

Maximum Likelihood Estimation for $\theta := \underline{\{W, \mu, \sigma^2\}}$

$$-\log p(X \mid \mu, w, \sigma^2) = -\sum_{i=1}^{n} \log p(x_i \mid W, \mu, \sigma^2)$$

(1)

$$= \frac{nd}{2} \log 2\pi + \frac{n}{2} \log |G| + \frac{1}{2} \sum_{i=1}^{n} (x_i - \mu)^T G^{-1} (x_i -$$

Optimal parameters :

- $\mu_{MLE} = \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$

$$(1) \Rightarrow -\log p(X \mid W, \mu, \sigma^2) = \frac{n}{2} \{ d \log 2\pi + \log |G| + \text{Tr}(G^{-1} S') \}$$

where $S' := \frac{1}{\,} \sum^{n} (x_i - \mu_{MLE})(x_i - \mu_{MLE})^T$

$n \quad i=1$

- $$W_{MLE} = U_m \left(L_m - \sigma^2 I\right)^{\frac{1}{2}} R,$$

  where $U_m$ is a $d \times q$ matrix whose columns are the eigenvectors $S^t$, $L_m$ is a diagonal matrix containing the eigenvalues of $S^t$, and $R$ an arbitrary orthogonal matrix. (typically taken as $U_m^T$).

- $$\sigma^2_{MLE} = \frac{1}{d-q} \sum_{i=q+1}^{d} \lambda_i$$