MODULE 1.2: DIMENSIONALITY REDUCTION (PRINCIPAL COMPONENT ANALYSIS)

EXPECTATION OF A RANDOM VARIABLE

Expectation of a random variable is an average of all realizations of X weighted by likelihoods

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \, p_X(x) \, dx$$

Intuition: Regular average: we sum all values and divide by number of values

Expectation: weight values by their relative likelihoods and sum up

In the multivariate case, $\mathbb{E}[X] = [\mathbb{E}(X_1), \mathbb{E}(X_2), \ldots, \mathbb{E}(X_N)]$

Side Note: You can show that for a Gaussian distribution

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/\sigma^2} \, dx = \mu$$

VARIANCE

Measure of variability around the mean weighted by likelihoods

$$\text{var}[X] = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \int (x - \mathbb{E}[X])^2 p_X(x) dx$$

In the multivariate case this idea becomes the variance-covariance matrix

$$\Sigma_1 = \mathbb{E}\left[(X - E[X])(X - E[X])^T\right]$$

We can think of the diagonal of Σ_1 as element-wise variance of the multivariate random variable

$$\Sigma_{ii} = var[X(i)] = \mathbb{E}\left[(X(i) - \mathbb{E}[X(i)])^2\right]$$

i.e measure of variability of the i^{th} entry.

Covariance is the term used to express the relationship (with respect to the mean) between components of the vector random variable.

$$\Sigma_{i,j} = \mathbb{E}\left[(X(i) - \mathbb{E}[X(i)])(X(j) - \mathbb{E}[X(j)])\right]$$

if $\Sigma_{ij} = 0$, components are unrelated : orthogonal

$\Sigma_{ij} > 0$ ($\Sigma_{ij} < 0$), they are positively (negatively) related

NOTE:

① If $\mathbb{E}[X] = \vec{0}$ then the covariance matrix for any realization $\vec{x}$ is

$$\Sigma = \mathbb{E}[\vec{x}, \vec{x}^T]$$

② Covariance matrix $\Sigma = \mathbb{E}\left[(\vec{x} - \vec{\mu})(\vec{x} - \vec{\mu})^T\right]$ where $\vec{\mu} = \mathbb{E}(\vec{x})$

7 3 6
6 5.5

is symmetric i.e $\Sigma_{ij} = \Sigma_{ji}$, and infact positive semi-definite (can you show this?)

DISCRETE RANDOM VARIABLES

Assuming there are a countable number N realizations of random variable X, the mean is given by

$$\mathbb{E}(X) = \sum_{i=1}^{N} P(X=\vec{x}_i)\,\vec{x}_i = \vec{\mu}$$

the covariance matrix is given by

$$\Sigma = \sum_{i=1}^{N} P(X=\vec{x}_i)(\vec{x}_i - \vec{\mu})(\vec{x}_i - \vec{\mu})^T$$

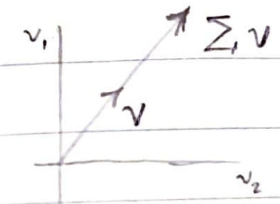
Assuming $P(X=\vec{x}_i) = 1/N \;\forall i$, what is $\vec{\mu}$ and Σ?

Eigenvalues & Eigenvectors of Covariance Matrix

Eigenvector: vector v such that

$$\Sigma v = \lambda v$$

where λ is the corresponding eigenvalue



Σv & v are collinear

If $\vec{v}$ is an eigenvector, $\alpha\vec{v}$ is also an eigenvector for any $\alpha \in \mathbb{R}$,

$$\Sigma(\alpha\vec{v}) = \alpha\Sigma\vec{v} = \alpha\lambda\vec{v} = \lambda(\alpha\vec{v}).$$

So we can typically normalize an eigenvector, without loss of generality by dividing by its norm

$$\vec{v} \leftarrow \vec{v}/\|\vec{v}\| \quad \text{for } \|\vec{v}\|^2 \neq 1.$$

NOTE

1. Because Σ is PSD (and symmetric), all its eigen values are real and nonnegative (Check that this is so!). This is also the case for complex-valued random vectors (for which Σ is now Hermitian)
2. In addition, the eigenvectors are orthogonal.

Proof (2): Assume $\vec{v}$ and $\vec{u}$ are eigenvectors associated with with eigen values λ and μ, respectively, and $\lambda \neq \mu$, then

$$\Sigma\vec{v} = \lambda\vec{v}, \qquad \Sigma u = \mu\vec{u}$$

Since the matrix Σ is Hermitian, we have that

$$\vec{u}^H\Sigma\vec{v} = \vec{u}^H\Sigma^H\vec{v} = (\Sigma\vec{u})^H\vec{v}$$

Now if we make $\Sigma\vec{v} = \lambda\vec{v}$ on the leftmost side, and $\Sigma\vec{u} = \mu\vec{u}$ on the rightmost side, we have

$$u^H \lambda v = \lambda u^H v = (\mu u)^H v = \bar{\mu} \vec{u}^H v$$

$$\lambda \vec{u}^H \vec{v} = \bar{\mu} \vec{u}^H \vec{v} \text{ only if } \vec{v}^H \vec{u} = 0 \text{ since } \mu \neq \lambda$$

and $\mu, \lambda \in \mathbb{R}$

Eigenvector Matrix

Let T be the eigenvector matrix whose columns are the ordered eigenvectors of Σ. Since the eigenvectors are orthonormal,

$$T = [\vec{v}_0 \ \vec{v}_1 \ \ldots \ \vec{v}_{N-1}]$$

$$T^H T = I \iff T \text{ is unitary}$$

PCA TRANSFORM

We define the Principal Component Analysis Transform as

$$\vec{y} = T^H \vec{x}$$ ⟸ The projection onto the coordinates $v_0, \ldots, v_{N-1}$ (or 'basis')

and the inverse, iPCA transform

$$\vec{\tilde{x}} = T \vec{y}$$ ⟸ scale the PC (or 'basis') by $y_0 \ldots y_{N-1}$ & add

Since T is unitary $\vec{\tilde{x}} = T\vec{y} = TT^H x = I\vec{x} = \vec{x}$

Thus y is an equivalent representation of x

and $\|\vec{x}\|^2 = \|\vec{y}\|^2$ $\Leftarrow$ The energy composition of the vector is conserved

The PCA transform is defined for any vector x of appropriate dimension but, in particular, we expect it to give suitable principal components when x is a realization of X

Do you recognize similarities with the DFT?

DFT

$$\hat{f}_k = \sum_{j=0}^{N-1} f_j \, \omega_N^{kj}$$

PCA

$$y_k = \sum_{j=0}^{n-1} x_j V_{kj}$$

Different 'bases', & more specific to a given random variable,

but the sense of taking to a new 'basis' persists.

Compression with PCA

Recall that for the DFT, we compress by retaining

$C < N$ DFT coefficients, when doing the iDFT

$$\tilde{f}_j = \sum_{k=0}^{C-1} \hat{f}_k \, \omega_N^{jk}$$

In a similar fashion we can compress using PCA by retaining $C < N$ PCA coefficients

$$\tilde{x}_j = \sum_{k=0}^{C-1} y_k \, v_{jk}$$

or

$$\tilde{x} = T \tilde{y}$$

where

$$\tilde{y}_k = y_k \quad \text{for } k < C, \quad \tilde{y}_k = 0 \text{ otherwise}$$