Basics of Probability, Random Processes and Linear Algebra

Pattern Recognition, Jul-Nov 2017

Indian Institute of Technology Madras

7 Aug- 11 Aug

Problem

A station attempts to transmit a packet. Let p be the probability that it will not collide. Find the expected number of collisions for successful packet transmission.

Ans:

X	0	1	2	3	...	n
p(X)	p	pq	q^2p	q^3p	...	q^np

This is an ordered series.

$$E[X] = \sum x p_x(x)$$

$$= \sum_{i=0}^{\infty} ipq^i$$

$$= 0 + pq + 2pq^2 + 3pq^3 + \ldots$$

$$= p[q + 2q^2 + 3q^3 + \ldots]$$

Let $S = q + 2q^2 + 3q^3 + \ldots$

$$Sq = q^2 + 2q^3 + 3q^4 + \ldots$$

$$S(1-q) = q + q^2 + q^3 + \ldots$$

$$S = \frac{q}{(1-q)^2}$$

Hence, $E[X] = \frac{q}{p}$

$$Var(X) = E[(X - \mu_x)^2]$$

$$= E[X^2] - (E[X])^2$$

$$= E[X^2] - (\frac{q}{p})^2$$

- Find $E[X^2]$.

  Ans: $E[X^2] = pq \sum_{k \geq 1}^{\infty} k^2 q^{k-1}$

  Hint: $\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}$
  Take derivative on both sides.

Cross co-variance

Cross co-variance

Cross-covariance of two random variables X and Y is given as:
$Cov(X, Y) = E[(X(t) - \mu_x)(Y(t) - \mu_y)]$

Correlation coefficient

Correlation coefficient for two random variables X and Y is given as:
$S_{XY} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}; \quad S_{XY} \leq 1$

Some properties

- $Cov(X, Y) = Cov(Y, X)$
- $X = Y \implies S_{XY} = 1$
- X and Y are independent $\implies E[X] = E[Y]$
- X and Y are independent $\implies$ $Cov(X, Y) = E[XY] - E[X]E[Y] = 0$

Some properties

- $E[aX] = aE[X]$
- $E[a_1X_1 + a_2X_2 + ...] + b = a_1E[X_1] + a_2E[X_2] + ..$
- $Var(aX) = a^2Var(X)$
- $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$
- X and Y are independent $\implies$ $Var(X + Y) = Var(X) + Var(Y)$

Random Vector

A d dimensional random vector X is represented as :

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix} = [x_1 \;\; x_2 \;\; \ldots \;\; x_d]^t$$

Joint pdf

$P(X_1 X_2 X_3 ... X_d) \epsilon A)] = \int_A f_X(x_1, x2, x_3, ...x_d) dx_1 dx_2 dx_3 ... dx_d$

$= \int_A f_X(x_1) dx_1 \int_A f_X x_2) dx_2 \int_A f_X(x_3) dx_3 ... \int_A f_X(x_d) dx_d,$

if x_1,x_2,x_3... are independent

Mean vector and Covariance matrix

Mean vector

$$E[\vec{X}] = \begin{bmatrix} E[\vec{X_1}] \\ E[\vec{X_2}] \\ \vdots \\ E[\vec{X_d}] \end{bmatrix}$$

Covariance matrix

$$Cov(\vec{X}) = E[(\vec{X} - E[\vec{X}])(\vec{X} - E[\vec{X}])^t]$$

$$Cov[\vec{X}] = \begin{bmatrix} Cov[X_1X_1] & Cov[X_1X_2] & \ldots Cov[X_1X_d] \\ Cov[X_2X_1] & Cov[X_2X_2] & \ldots Cov[X_2X_d] \\ \vdots & & \\ Cov[X_dX_1] & Cov[X_dX_2] & \ldots Cov[X_dX_d] \end{bmatrix}$$

- $\vec{Y} = A\vec{X} + B \implies E[\vec{Y}] = AE[\vec{X}] + \vec{B}$
- $Cov(\vec{Y}) = ACov(\vec{X})A^t$
- Covariance is estimated from the data as:

$$Cov(X_1, X_2) = E[(X_1 - \mu_X 1)(X_2 - \mu_X 2)] \tag{1}$$

$$= \frac{1}{N_1 - 1} \sum_{i=1}^{N_1} (x_1 i - \mu x_1)^2$$

(The denominator term is $N_1 - 1$ because N^{th} value is deterministic if μ_x and $N_1 - 1$ values are known.)

Properties of vectors

- d dimensional vector x is represented as :

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix} = [x_1 \ x_2 \ \ldots \ x_d]^t$$

- Euclidean distance with respect to origin :

$$\|\vec{x}\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_d^2}$$

- Cauchy Shwartz Inequality :

$$\|\vec{x} + \vec{y}\| \leq \|\vec{x}\| + \|\vec{y}\| \quad \text{(Proof: Triangle law of vector addition)}$$
$$\|\vec{x}^t \vec{y}\| \leq \|\vec{x}\| \|\vec{y}\|$$

Proof : $\vec{x}^t \vec{y}$ is the inner product which is equal to $\|\vec{x}\| \|\vec{y}\| \cos\theta$ and $\cos\theta \leq 1$.

- Difference between 2 vectors :

$$\vec{x_1} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \vec{x_2} = \begin{bmatrix} 4 \\ 5 \end{bmatrix} \Rightarrow \vec{e} = (\vec{x_1} - \vec{x_2}) = \begin{bmatrix} -2 \\ -2 \end{bmatrix} \text{ and } \|\vec{e}\| = 2\sqrt{2}$$

- Inner Product :

$$\vec{x}^t \vec{y} = \begin{bmatrix} a_1 a_2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} a_1 b_1 + a_2 b_2 \end{bmatrix} \quad \text{(Scalar)}$$

- Cosine Similarity :

$$\cos\theta = \frac{\vec{x_1}^t \vec{x_2}}{\|\vec{x_1}\| \|\vec{x_2}\|}$$

But we cannot simply use it everywhere. The cosine similarity between *This is a PR class* and *This is not a PR class* is very high. (Converting each sentence to a binary vector denoting if a word is present in the sentence or not, we get the vectors as [1 1 0 1 1 1] and [1 1 1 1 1 1], having cosine sim. $= \frac{5}{\sqrt{30}} \approx 1$)

Hence problem : The cosine similarity showed that the sentences are highly similar but we can see that semantically they are completely opposite. Solution : We consider word co-occuring probability.

- Outer Product :

$$\vec{x}\vec{y}^t = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \begin{bmatrix} b_1 b_2 \end{bmatrix} = \begin{bmatrix} a_1 b_1 & a_1 b_2 \\ a_2 b_1 & a_2 b_2 \end{bmatrix} \quad \text{(Not scalar)}$$

- $\vec{x}^t M \vec{x}$ is scalar $\qquad (\because \vec{x}^t_{1xn} \quad M_{nxn} \quad \vec{x}_{nx1})$

  Notice that it is a quadratic equation. Imagine it by putting M=[1], so $\vec{x}^t\vec{x}$=s (some scalar). Now if $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ then using inner product property, $\vec{x}^t\vec{x} = x_1^2 + x_2^2 =$ s which is quadratic.

- Gradient of a function with respect to a vector :

$$\nabla f(\vec{x}) = \frac{\partial f(\vec{x})}{\partial x} = \begin{bmatrix} \frac{\partial f(\vec{x})}{\partial x_1} \\ \frac{\partial f(\vec{x})}{\partial x_2} \\ \vdots \\ \frac{\partial f(\vec{x})}{\partial x_d} \end{bmatrix}$$

Here $f(\vec{x})$ is scalar while $\vec{x}$ is a vector. Note that scalar to vector differentiation is a vector.

Example : $f(\vec{x}) = 2x_1^2 x_2 + 3x_1 x_2^3 - 5x_1 + 2x_2 + 6$ and $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

$$\implies \nabla f(\vec{x}) = \begin{bmatrix} 4x_1 x_2 + 3x_2^3 - 5 \\ 2x_1^2 + 9x_1 x_2^2 + 2 \end{bmatrix}$$

- Jacobian (matrix of derivatives) :
  It is used when we change variables, say from d dimensional $\vec{x}$ to n dimensional $\vec{y}$

$$y_1 = f_1(x_1, x_2, \ldots, x_d)$$
$$y_2 = f_2(x_1, x_2, \ldots, x_d)$$
$$y_n = f_n(x_1, x_2, \ldots, x_d)$$

$$\text{Jacobian} = \mathbf{J} = \begin{bmatrix} \nabla^t f_1(\vec{x}) \\ \nabla^t f_2(\vec{x}) \\ \vdots \\ \nabla^t f_n(\vec{x}) \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_d} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_d} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_d} \end{bmatrix}$$

  Basically, the result of the previous property got transposed and became a row of the jacobian matrix for each of the n functions.

- Eigenvalue and Eigenvector
  $A\vec{e} = \lambda\vec{e}$, where λ is the eigenvalue and $\vec{e}$ is the eigenvector.

  Are eigenvectors unique? : No, we can multiply any eigenvector with a scalar and that scaled vector will also be an eigenvector.

  Are eigenvalues unique? : No, the characteristic equation ($|A - \lambda I| = 0$) whose roots are the eigenvalues, can have repeated roots.

  $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ The multiplicity of eigenvalue is 3 but still we can get linearly independent eigenvectors [0 0 1] , [1 0 0] , [0 1 0] for this matrix.

  Read more properties of eigenvalue eigenvectors yourself

Calculate the eigenvalues and eigenvectors for the matrix A :

$$\begin{bmatrix} 1 & 2 & 1 \\ 6 & -1 & 0 \\ -1 & -2 & -1 \end{bmatrix}$$

We get λ = 0, 3, -4 on solving

$$|A - \lambda I| = 0 \implies \begin{vmatrix} 1-\lambda & 2 & 1 \\ 6 & -1-\lambda & 0 \\ -1 & -2 & -1-\lambda \end{vmatrix} = 0$$

Now, to find the eigenvectors, let us first take $\lambda = 0$ and put in $A\vec{e} = \lambda\vec{e} \implies A\vec{e} = 0$

$$\begin{bmatrix} 1 & 2 & 1 \\ 6 & -1 & 0 \\ -1 & -2 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = 0 \implies \begin{matrix} x + 2y + z = 0 \\ 6x - y = 0 \\ -x - 2y - z = 0 \end{matrix}$$

Solving we get $x = c, y = 6c, z = -13c$ where c can take any value and hence multiple eigenvectors.

Similarly find eigenvectors for $\lambda = 3$ and -4

Eigenvalue decomposition (EVD)

Consider a matrix of size d. There can be several vectors $\vec{e_i}$ satisfying the following

$$A\vec{e_1} = \lambda_1 \vec{e_1}$$
$$A\vec{e_2} = \lambda_2 \vec{e_2}$$
$$\vdots$$
$$A\vec{e_d} = \lambda_d \vec{e_d}$$

In matrix form,

$$AX = X\Lambda$$

A can also be written as

$$A = X\Lambda X^{-1}$$

where,

EVD contd ...

$$X = [\vec{e_1} \ \vec{e_2} \ \cdots \ \vec{e_d}]$$

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & 0 & \cdots & \vec{e_d} \end{bmatrix}$$

- Every column in matrix X represents an Eigenvectors (direction).
- The magnitude of diagonal values represent the strength of the corresponding Eigen direction.
- If A is symmetric positive definite i.e., $X^{-1} = X^t$.

$$\hat{A} = X \Lambda X^{-1} \text{ then,}$$

1. Eigenvectors are orthogonal
2. Eigenvalues are positive

Singular Value Decomposition (SVD)

If the matrix A is not a square matrix, it can be decomposed as

$$A = U\Sigma V^t$$

where U and V matrices contain left and right singular vectors

$$\begin{aligned} AA^t &= U\Sigma V^t (U\Sigma V^t)^t \\ &= U\Sigma V^t V\Sigma U^t \\ &= U\Sigma^2 V^t \end{aligned}$$

- U,V are unitary matrices and Σ is a diagonal matrix.
- U and V matrices also contain Eigenvectors of AA^t and A^tA respectively.
- *Orthogonality is guaranteed in SVD as opposed to EVD.*