Machine Learning - HW 1

Abhilekh Pandey, Nurgun Rafizade, Vlada Petrusenko

February 18, 2020

1 Problem 1

We consider vector space V as the set of all polynomials with degree less than m

1.a)

Take:

$\vec{v_1} = a_1x^0 + a_2x^1 + a_3x^2 + a_4x^3 + ... + a_mx^{m-1}$

$\vec{v_2} = b_1x^0 + b_2x^1 + b_3x^2 + b_4x^3 + ... + b_mx^{m-1}$

Then:

$\vec{v_1} + \vec{v_2} = (a_1 + b_1)x^0 + (a_2 + b_2)x^1 + (a_3 + b_3)x^2 + ... + (a_m + b_m)x^{m-1}$

Thus:

$\vec{v_1} + \vec{v_2} \in V$

Associativity and Commutativity follow from polynomial addition.

$-\vec{v_1} = -a_1x^0 - a_2x^1 - a_3x^2 - a_4x^3 - ... - a_mx^{m-1}$

Thus:
$-\vec{v_1} \in V$

$0 \in V$ follows

Distibutativity of constant follows from polynomial addition.

V satisfies all the properties of a vector space.

1.b)

Take $m = 3$ for V. Take B as the standard basis $\{1, x, x^2\}$

Take A as the set of polynomials $\{1, 2x, x^2 + 2x + 1\}$

The Matrix of A can be given by:

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

If A is a basis of vector space V, then there exists no non trivial solution for:

$A\vec{v} = 0$

Matrix A is already in an upper triangle echelon form. Thus there is no non trivial solution for the equation. A is a basis of V

1.c)

$F : R^2 \to P_2, F(\vec{c}) = c_1 x + c_2$

Take:

$$\vec{x} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}$$

$$\vec{y} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$$

Checking for linearity:

$$F(\vec{x} + \vec{y}) = F \begin{bmatrix} \alpha_1 + \beta_1 \\ \alpha_2 + \beta_2 \end{bmatrix}$$

$F(\vec{x} + \vec{y}) = (\alpha_1 + \beta_1)x + (\alpha_2 + \beta_2)$

$F(\vec{x} + \vec{y}) = (\alpha_1 x + \alpha_2) + (\beta_1 x + \beta_2)$

$F(\vec{x} + \vec{y}) = F(\vec{x}) + F(\vec{y})$

$$F(c \cdot \vec{x}) = F \begin{bmatrix} c \cdot \alpha_1 \\ c \cdot \alpha_2 \end{bmatrix}$$

$F(c \cdot \vec{x}) = (c \cdot \alpha_1)x + (c \cdot \alpha_2)$

$F(c \cdot \vec{x}) = c \cdot F(\vec{x})$

F is a linear in C.

1.d)

It is given for V that:

$$< f(x), g(x) >= \int_0^1 f(x)g(x)dx$$

We have basis A $= \{1, x, x^2\}$

We form a orthonormal basis around x.

$\vec{a} = 1$

$\vec{b} = x$

$\vec{c} = x^2$

We have the projection of $\vec{a}$ on $\vec{b}$ defined as:

$$proj_{\vec{b}}(\vec{a}) = \frac{< \vec{a}, \vec{b} >}{< \vec{b}, \vec{b} >} \cdot \vec{b}$$

Orthonormal $\vec{a} = \frac{\vec{a} - proj_{\vec{b}}(\vec{a})}{||\vec{a}||}$

Orthnormal $\vec{a} = -\sqrt{3}x$

Orthonormal $\vec{c} = \frac{\vec{c} - proj_{\vec{b}}(\vec{c})}{||\vec{c}||}$

Orthonormal $\vec{a} = -3\sqrt{5}x + \frac{1}{4\sqrt{5}}x^2$

Orthonormal $\vec{b} = \frac{\vec{b}}{||\vec{b}||}$

Orthonormal $\vec{b} = -\frac{1}{\sqrt{3}}x$

The orthonormal basis is $\{-\sqrt{3}x, -\frac{1}{\sqrt{3}}x, -3\sqrt{5}x + \frac{1}{4\sqrt{5}}x^2\}$

2 Problem 2

2.I

We have a sample space of $\Omega = \{\text{All combinations of 3 consecutive coin tosses}\}$.

2.I.a)

Event A is defined to be {HHH}

Probability of A $= \frac{1}{8}$

2.I.b)

Event B is defined to be {HTT, THT, TTH}

Probability of B $= \frac{3}{8}$

2.I.c)

Let H denote the number of heads in a sample point in Ω

Probability of at least one head can be given by:

$P(H \geq 1) = P(H = 1) + P(H = 2) + P(H = 3)$

$P(H \geq 1) = \frac{3}{8} + \frac{3}{8} + \frac{1}{8}$

$P(H \geq 1) = \frac{7}{8}$

Probability of at least two heads can be given by:

$P(H \geq 2) = P(H = 2) + P(H = 3)$

$P(H \geq 2) = \frac{3}{8} + \frac{1}{8}$

$P(H \geq 2) = \frac{4}{8}$

The intersection of events at least one head and at least two heads is the event at least two heads.

Thus the probability of observing two heads given at least one head is observed is:

$P(H \geq 2 | H \geq 1) = \frac{P(H \geq 2)}{P(H \geq 1)}$

$P(H \geq 2 | H \geq 1) = \frac{4/8}{7/8}$

$P(H \geq 2 | H \geq 1) = \frac{4}{7}$

2.II

A dice is rolled twice. Take X and Y as the result of the first and second roll respectively.

2.II.a)

$P(X = 2, Y = 6) = \frac{1}{6} \cdot \frac{1}{6}$

$P(X = 2, Y = 6) = \frac{1}{36}$

2.II.b)

$P(X > 3) = P(X = 4) + P(X = 5) + P(X = 6)$

$P(X > 3) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6}$

$P(X > 3) = \frac{1}{2}$

$P(Y = 2) = \frac{1}{6}$

Since X and Y are independent of each other,

$P(X > 3|Y = 2) = P(X > 3)$

$P(X > 3|Y = 2) = \frac{1}{2}$

3 Multivariate Random Variables Problem

3.I

3.I.a)

Marginals:

$\rho_X(0) = 0.1 + 0.2 + 0.2 = 0.5$

$\rho_X(1) = 0.3 + 0.1 + 0.1 = 0.5$

$\rho_Y(0) = 0.1 + 0.3 = 0.4$

$\rho_Y(1) = 0.2 + 0.1 = 0.3$

$\rho_Y(2) = 0.2 + 0.1 = 0.3$

3.I.b)

For X:

$E(X) = \sum_X x\rho(x)$

$E(X) = P(X = 0) \cdot 0 + P(X = 1) \cdot 1$

$E(X) = (0.1 + 0.2 + 0.2) \cdot 0 + (0.3 + 0.1 + 0.1) \cdot 1$

$E(X) = 0.5 \cdot 0 + 0.5 \cdot 1$

$E(X) = 0.5$

For Y:

$E(Y) = \sum_Y y\rho(y)$

$E(Y) = p(Y = 0) \cdot 0 + p(Y = 1) \cdot 1 + p(Y = 2) \cdot 2$

$E(Y) = (0.1 + 0.3) \cdot 0 + (0.2 + 0.1) \cdot 1 + (0.2 + 0.1) \cdot 2$

$E(Y) = 0.4 \cdot 0 + 0.3 \cdot 1 + 0.3 \cdot 2$

$E(Y) = 0.9$

3.I.c)

$E(Y|X = 0) = \sum_y y \cdot \rho(y|x)$

$E(Y|X = 0) = \frac{\sum_y y \cdot \rho(0,y)}{\rho_X(0)}$

$E(Y|X = 0) = \frac{0+1 \cdot \rho(0,1)+2 \cdot \rho(0,2)}{0.5}$

$E(Y|X = 0) = \frac{0.6}{0.5}$

$E(Y|X = 0) = 1.2$

3.II

Analyzing set C, the elements of the set are: (0,0), (0,1), (0,2), (0,-1), (0,-2), (1,0), (1,1), (1,-1), (-1,0), (-1,1), (-1,-1). In total 11 elements, All of them are equally probable to be picked.

$X\backslash Y$	-2	-1	0	1	2
-1	0	1/11	1/11	1/11	0
0	1/11	1/11	1/11	1/11	1/11
1	0	1/11	1/11	1/11	0

3.II.a)

The joint PMF of each of these elements is $\frac{1}{11}$, as all of them are equally probable.
For marginal PMFs (all of them follow just from definition by calculating sum of respective joint probabilities):

$p_X(0) = \frac{5}{11}$

$p_X(1) = \frac{3}{11}$

$p_X(-1) = \frac{3}{11}$

$p_Y(0) = \frac{3}{11}$

$p_Y(1) = \frac{3}{11}$

$p_Y(-1) = \frac{3}{11}$

$p_Y(2) = \frac{1}{11}$

$p_Y(-2) = \frac{1}{11}$

3.II.b)

$\rho(X|Y = 1) = \frac{\rho(x,1)}{\rho_y(1)}$

$\rho(-1|Y = 1) = \frac{1}{3}$

$\rho(0|Y = 1) = \frac{1}{3}$

$\rho(1|Y = 1) = \frac{1}{3}$

3.II.c)

No. Otherwise $P(X \cap Y) = P(X)P(Y)$ would be true but it's not.