MACHINE LEARNING

Materi#2 : Fondasi Matematika

Deskripsi

Tujuan dari pembelajaran Materi ke-2 adalah; Mahasiswa akan mendapatkan pengetahuan dasar matematika yang digunakan pada Machine Learning. Mahasiswa akan diberikan kasus – kasus dan harus diselesaikan dengan menggunakan pendekatan matematika (CPMK1)

Capaian Pembelajaran

Pada pertemuan ini, kita akan mempelajari:

- Probabilitas
- Probability Density Function
- Expectation and Variance
- Bayesian Probability
- Gaussian Distribution
- Teori Keputusan
- Matriks

Referensi

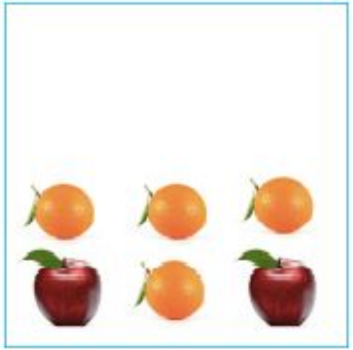
- Jan Wira Gotama Putra, Pengenalan Konsep Pembelajaran Mesin dan Deep Learning, Tokyo, Jepang: -, 2020.
- Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.
- YouTube 'Statistical Inference Prof Brian Caffo'

Agenda

- Memahami fondasi matematis dalam Machine Learning

Probabilitas

- Didunia banyak hal yang tidak pasti (uncertain)
- Machine Learning berurusan dengan ketidakpastian (Uncertainty)



$$P(E) = \frac{X}{N}$$

P : Probabilitas

E : kejadian yang diinginkan

X : banyaknya cara kejadian akan terjadi

N : Total kejadian yang mungkin terjadi

- Terdapat 2 buah kotak dan merah. Pada kotak biru terdapat 2 apel dan 3 jeruk; pada kotak merah terdapat 3 apel dan 1 jeruk
- Kotak disebut *random variabel* (*k*, melambangkan kotak; nilainya merah atau biru)
- Buah dilambangkan dengan variabel *b;* nilainya apel dan jeruk
- Nilai probabilitas lebih besar dari 0 dan kurang dari 1 (0 <= P <= 1)

Contoh sederhana: Kita ingin mengambil buah dari salah satu kotak

1. Berapa peluang untuk memilih apel dan jeruk pada kotak biru;
   P(b = apel) = 2/ 6
   P(b = jeruk) = 4/ 6

Tugas Probabilitas

1. Berapa nilai probabilitas mengambil sebuah apel pada kotak biru atau jeruk pada kotak merah?
2. Berapa probabilitas terambilnya apel merah pada kotak mana saja?

Petunjuk: Silahkan baca referensi lain terkait probabilitas!

Probability Density Function/Fungsi Kepadatan Probabilitas

Perusahaan jasa penjualan telur ayam kampung yang dikelola sendiri oleh Pak Hadi, mempunyai 3 orang karyawan. Setiap bulannya pak Hadi membayar upah setiap karyawannya sebesar 1 juta rupiah, dia sendiri setiap bulannya mengambil bayaran sebesar 9 juta rupiah. Kemudian dia mengatakan bahwa rata-rata upah dalam perusahaannya adalah 3 juta rupiah

Perhitungan statistik

Hadi	**Rp. 9 jt**
Karyawan 1	**Rp. 1 jt**
Karyawan 2	**Rp. 1 jt**
Karyawan 3	**Rp. 1 jt**

Rata2 = Rp. 12 jt /4 orang
= Rp. 3 jt/orang

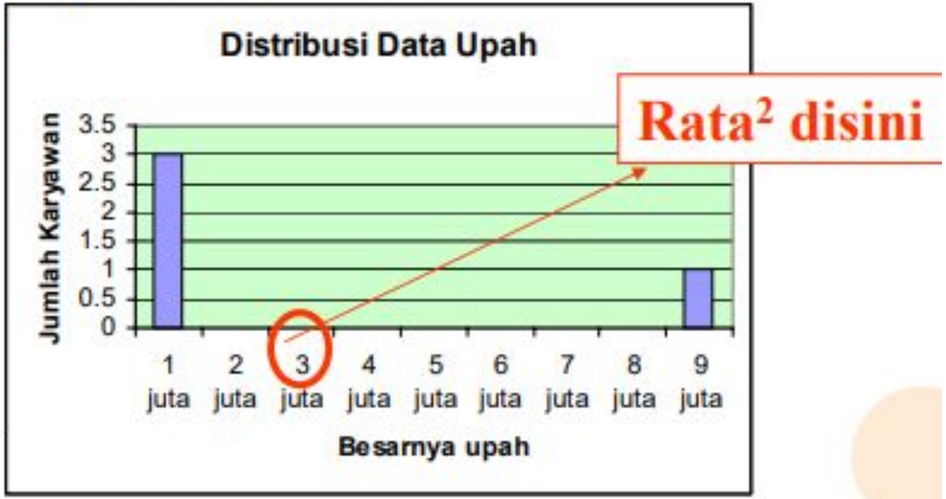
Apakah ini masuk akal ?

Kalo iya, karyawan yang mana yang mendapat upah 3 jt rupiah?

Kenyataannya tidak ada seorang karyawanpun yang mendapat upah 3 juta

Coba kita perhatikan masalah ini lebih jauh dengan mengamati distribusi data dari upah karyawan, dimana:
1 orang mendapat bayaran 9 juta rupiah, dan 3 orang mendapat upah 1 juta rupiah, sehingga dapat kita gambarkan distribusi data besarnya upah sebagai berikut:

Besarnya upah	Jumlah karyawan
1 juta	3
2 juta	0
3 juta	0
4 juta	0
5 juta	0
6 juta	0
7 juta	0
8 juta	0
9 juta	1



Distribusi data seperti ini adalah suatu kejadian dimana nilai rata-rata (statistik parametrik) tidak dapat menunjukkan hasil yang dapat menggambarkan kenyataan sesungguhnya.

Data hasil ujian pemrograman dari 20 mahasiswa Jurusan TI adalah seperti tabel di sebelah kanan. Dapat dinyatakan bahwa nilai rata-rata programming adalah 69.

Kesimpulan:

Pemrograman mahasiswa TI rata-rata lemah, karena tidak mencapai nilai 75 sebagai standard yang sudah ditentukan sebelumnya.

Kenyataannya adalah:

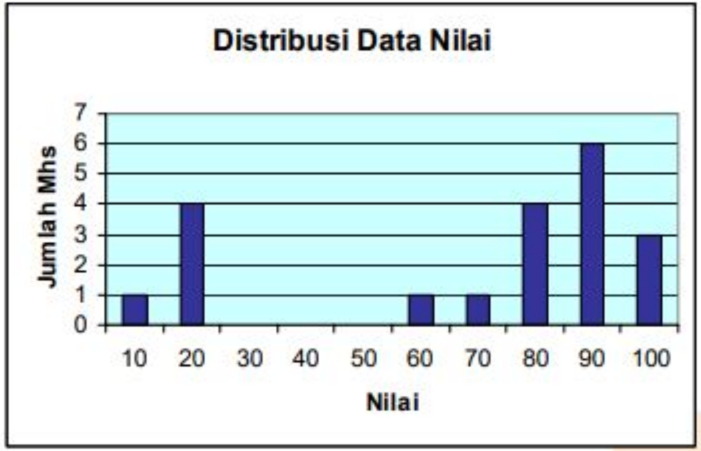
Hanya beberapa orang yang lemah, sedang sebagian besar (13 mhs) yang nilainya di atas 75.

no.mhs	Nilai
1	90
2	70
3	20
4	80
5	10
6	100
7	60
8	80
9	90
10	80
11	20
12	80
13	100
14	20
15	90
16	20
17	100
18	90
19	90
20	90

no.mhs	Nilai
1	90
2	70
3	20
4	80
5	10
6	100
7	60
8	80
9	90
10	80
11	20
12	80
13	100
14	20
15	90
16	20
17	100
18	90
19	90
20	90

Coba kita perhatikan distribusi nilai pemrograman dari mahasiswa TI ini.

Nilai	Jumlah Mhs
10	1
20	4
30	0
40	0
50	0
60	1
70	1
80	4
90	6
100	3



Dengan distribusi data ini, nilai rata-rata tidak dapat menunjukkan keadaan sebenarnya.

Bagaimana cara menunjukkan distribusi data agar kesimpulan yang diambil dapat menunjukkan keadaan sesungguhnya ?

Fungsi Kepadatan Probabilitas

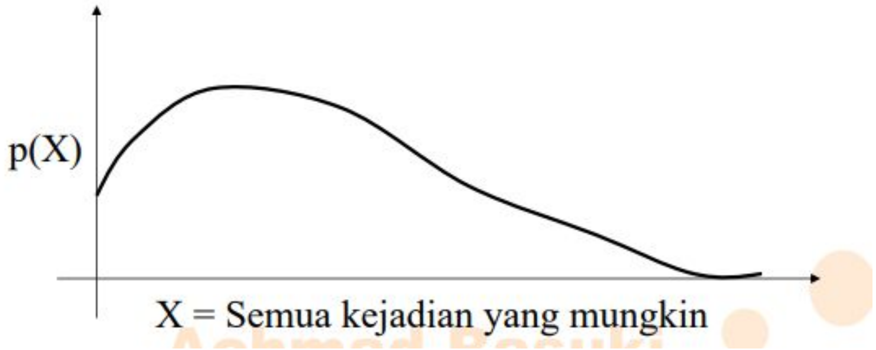
Definisi

- Fungsi kepadatan probabilitas atau *probability density function* (pdf) menyatakan nilai probabilitas dari setiap kejadian X dan dituliskan dengan p(X)
- Karena p(X) menyatakan nilai probabilitas maka $0 \leq p(X) \leq 1$
- Untuk semua kejadian maka jumlah nilai probabilitasnya adalah satu atau dituliskan dengan:

$$\sum_n p(X = x_n) = 1$$

Grafik

- Grafik yang menyatakan nilai kemungkinan dari setiap kejadian.
- Absis menyatakan kejadian yang mungkin.
- Ordinat menyatakan nilai kemungkinan $p(x_i)$



Kontinu VS Diskrit

Kontinu vs Diskrit

Pada dasarnya fungsi-fungsi di dalam statistik berdasarkan sifat kejadiannya dibedakan menjadi dua macam yaitu kontinu dan diskrit.

- Kontinu: kejadian yang mungkin jumlahnya tak berhingga dan operasionalnya dilakukan dalam bentuk kalkulus, misalkan untuk menghitung jumlah peluang semua kejadian dituliskan dengan:

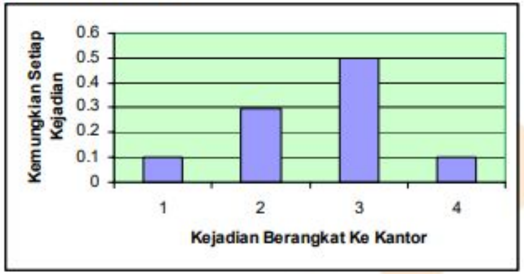
$$\int_{\forall x} f(x)dx = 1$$

- Diskrit: kejadian yang mungkin jumlahnya berhingga dan dapat berarti dilakukan secara berkala, operasionalnya menggunakan operasional fungsi diskrit, misalkan untuk menghitung jumlah peluang semua kejadian dituliskan dengan:

$$\sum_{n} p(X = x_n) = 1$$

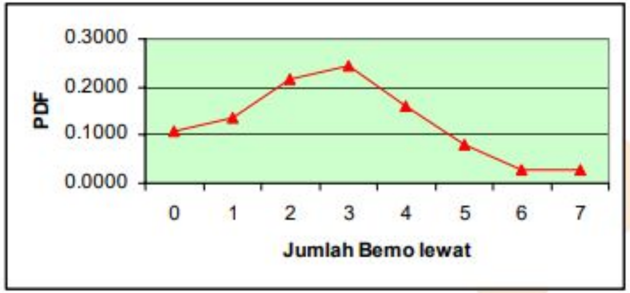
Contoh 1

- X adalah suatu kejadian seseorang akan berangkat ke kantor: kemungkinan dia berangkat naik mobil adalah 0.1, kemungkinan naik kendaraan umum 0.3, kemungkinan naik sepeda motor 0.5 dan kemungkinan tidak berangkat 0.1
- Fungsi kepadatan probabilitas dinyatakan dengan: $p(x_1)=0.1$, $p(x_2)=0.3$, $p(x_3)=0.5$ dan $p(x_4)=0.1$ dimana $X=\{x_1,x_2,x_3,x_4\}$ menyatakan kejadian-kejadian yang mungkin.
- Nilai probabilitas dari semua kemungkinan adalah $0.1+0.3+0.5+0.1 = 1$



Contoh 2

- Dari hasil pencatatan jumlah mobil bemo yang lewat setiap setengah jam di depan ITS diperoleh: tidak ada yang lewat: 4 kali, 1 bemo lewat: 5 kali, 2 bemo lewat: 8 kali, 3 bemo lewat: 9 kali, 4 bemo lewat: 6 kali, 5 bemo lewat: 3 kali, 6 bemo lewat: 1 kali, 7 bemo lewat: 1 kali.
- Absis (X) menyatakan jumlah bemo lewat dalam setengah jam
- Ordinat (Y) menyatakan kemunculan atau frekwensi kejadian dibagi dengan jumlah seluruh kejadian (37)



Histogram

- Histogram adalah suatu teknik untuk menyatakan jumlah munculnya setiap kejadian dari semua kejadian yang muncul.
- $H(x_n)$ menyatakan jumlah munculnya kejadian x_n.
- Fungsi kepadatan probabilitas $p(x_n)$ dapat dinyataka

$$p(x_i) = \frac{H(x_i)}{\sum_{j=1}^{n} H(x_j)}$$

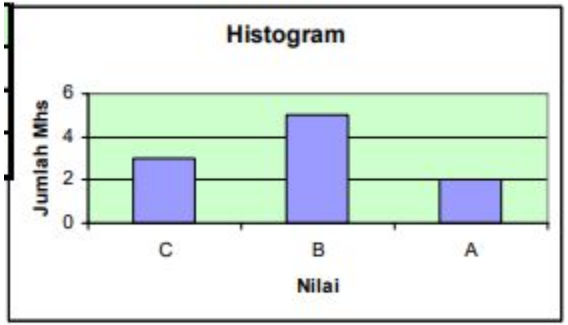
Data nilai test dasar pemrograman yang diikuti oleh 10 orang mahasiswa adalah sebagai berikut:

no. mhs	nilai
1	A
2	B
3	B
4	B
5	A
6	B
7	C
8	C
9	B
10	C

Berdasarkan nilai yang diperoleh dapat dinyatakan bahwa yang mendapat nilai C sebanyak 3 orang, yang mendapat nilai B sebanyak 5 orang dan yang mendapat nilai A sebanyak 2 orang. Nilai-nilai kemunculan ini disebut dengan histogram dari nilai ujian.

Nilai	Jumlah
C	3
B	5
A	2

Nilai	pdf
C	0.3
B	0.5
A	0.2



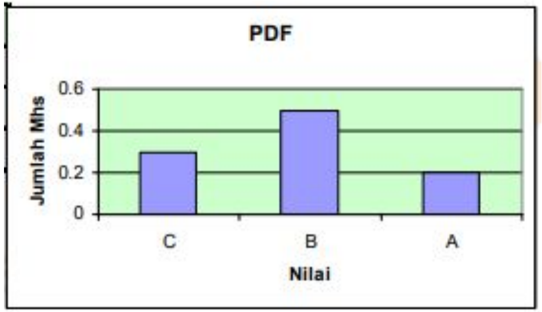


Diagram Pareto

- Suatu diagram yang digunakan untuk mencatat kemunculan setiap kejadian.
- Model diagram Pareto ini banyak digunakan untuk pencatatan kerusakan produksi.
- Diagram Pareto ini dapat juga disebut dengan diagram counting.
- Diagram Pareto ini banyak digunakan untuk menghasilkan nilai histogram dari suatu kejadian secara manual.

Pencatatan Kerusakan Produksi Tempe Setiap 1000 Bungkus Tempe Standard

Jenis Kerusakan	Counting	Jumlah
Kurang Masak	𝍸 𝍸 𝍸 𝍸 III	24
Kedelai Terlalu Busuk	𝍸 𝍸 𝍸 I	16
Kedelai Hancur	𝍸 𝍸	10
Tempe Pecah	𝍸 III	8

Histogram dan pdf dari kejadian di atas adalah:

Jenis Kerusakan	Histogram	PDF
Kurang Masak	24	0.4138
Kedelai Terlalu Busuk	16	0.2759
Kedelai Hancur	10	0.1724
Tempe Pecah	8	0.1379

Distribusi Frekuensi

- Distribusi frekwensi adalah suatu model perhitungan histogram dengan menggunakan pengelompokan data.
- Satu kelompok dapat dinyatakan sebagai satu range nilai dengan nilai tengah dianggap sebagai nilai yang mewakili kelompok tersebut.
- Kemunculan suatu kelompok dinamakan dengan frekwensi.

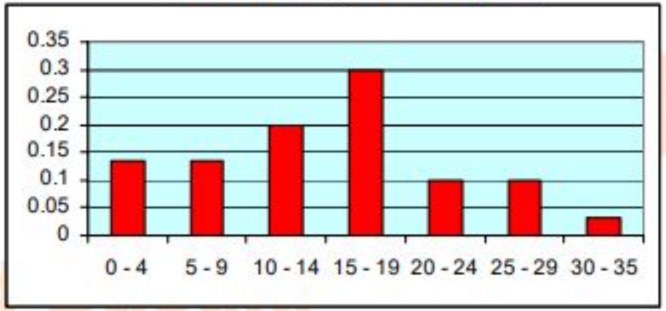
Data penjualan telor kampung setiap harinya pada toko MAJU MAKMUR dicatat selama 30 hari adalah sebagai berikut:

30	25	18	15	21	12	0	15	6	12
0	10	15	24	6	18	27	12	0	15
12	15	20	3	9	25	12	15	6	15

Distribusi frekwensi dengan range 5 adalah sebagai berikut:

Range	Median	Frekwensi
0 - 4	2	4
5 - 9	7	4
10 - 14	12	6
15 - 19	17	9
20 - 24	22	3
25 - 29	27	3
30 - 35	32	1

PDF dapat dihitung dengan frekwensi dibagi dengan jumlah seluruh kejadian (30)



Fungsi Kepadatan Kumulatif

- Fungsi Kepadatan Kumulatif atau Cumulative Density Function (CDF) adalah fungsi yang menjumlahkan nilai kemungkinan sampai suatu kejadian tertentu. Atau dituliskan dengan $p(X \leq x_i)$
- Bila X={x1, x2, x3, …, xn}, maka fungsi kepadatan kumulatif untuk $X=x_k$ dituliskan dengan:

$$p(X \leq x_k) = p(x_1) + p(x_2) + \ldots + p(x_k)$$

atau

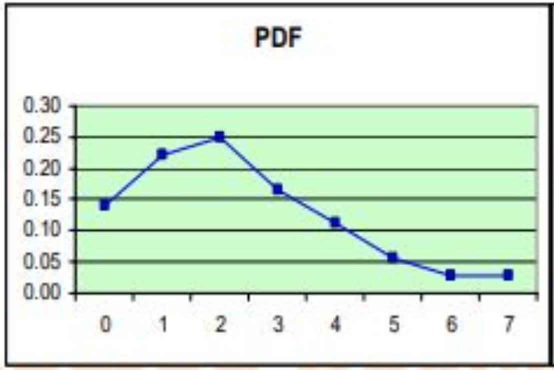
$$p(X \leq x_k) = \sum_{i=1}^{k} p(x_i)$$

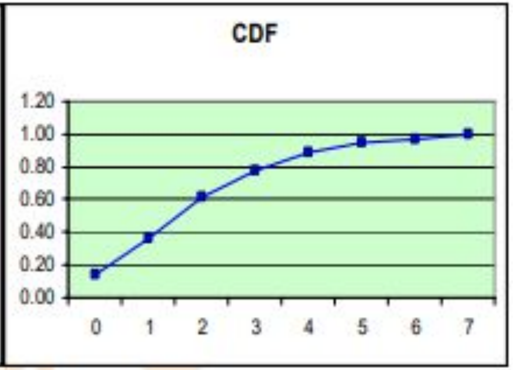
Diketahui frekwensi jumlah pelanggan yang melalui pintu kasir untuk setiap 5 menit sebuah supermarket adalah sebagai berikut:

Jumlah Plg	Frekwensi
0	5
1	8
2	9
3	6
4	4
5	2
6	1
7	1

Perhitungan PDF dan CDF adalah sebagai berikut:

Jumlah Plg	Frekwensi	PDF	CDF
0	5	5/36 = 0.14	0.14
1	8	8/36 = 0.22	0.14+0.22 = 0.36
2	9	9/36 = 0.25	0.36+0.25 = 0.61
3	6	6/36 = 0.17	0.61+0.17 = 0.78
4	4	4/36 = 0.11	0.78+0.11 = 0.89
5	2	3/26 = 0.06	0.89+0.06 = 0.94
6	1	1/36 = 0.03	0.94+0.03 = 0.97
7	1	1/36 = 0.03	0.97+0.03 = 1.00





Contoh Aplikasi - Pengamatan Nilai Matematika Mahasiswa TI

Nilai matematika 2 dari 30 mahasiswa Jurusan TI (kelas 2 TI-a) adalah sebagai berikut:

no.mhs	nilai	no.mhs	nilai	no.mhs	nilai
1	B	11	C	21	C
2	C	12	C	22	B
3	C	13	A	23	A
4	B	14	B	24	D
5	A	15	C	25	C
6	C	16	B	26	B
7	B	17	B	27	B
8	C	18	C	28	B
9	D	19	B	29	A
10	B	20	B	30	C

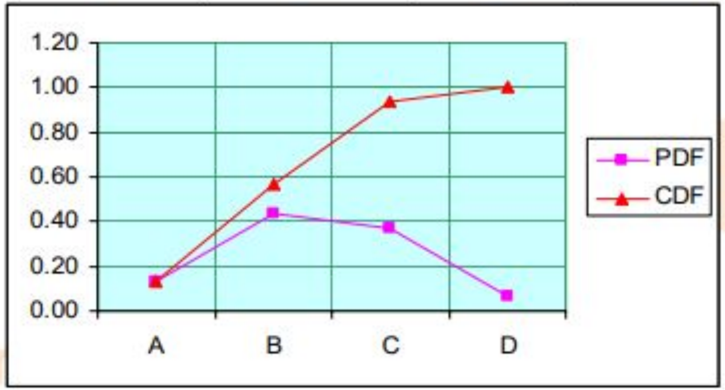
Nyatakan Histogram, PDF dan CDF dari data nilai mahasiswa di atas

Diagram distribusi frekwensi dari data nilai matematika tersebut adalah

Nilai	Jumlah mhs yang mendapat nilai
A	4
B	13
C	11
D	2

Histogram, PDF dan CDF diperoleh sebagai berikut:

Nilai	Histogram	PDF	CDF
A	4	0.13	0.13
B	13	0.43	0.57
C	11	0.37	0.93
D	2	0.07	1.00



Tugas PDF

Anda lakukan survey terhadap 20 orang teman anda yang dipilih secara acak. Tanyakan jenis acara TV yang sering ditonton oleh mereka dari acara-acara TV berikut ini:

(1) Olahraga
(2) Info Selebriti
(3) Berita
(4) Horor dan Misteri
(5) Film
(6) Film Kartun
(7) Komedi
(8) Sinetron

Buatlah Histogram, PDF dan CDF dari hasil survey tersebut, dan jangan lupa sebutkan segmen mahasiswa yang anda pilih berdasarkan jenis kelamin (berapa laki2 dan berapa wanita).

Expectation and Variance

The probability distribution of a discrete random variable X:

x	0	1	2	
$p(x)$	0.16	0.48	0.36	

← Values of X

← Probabilities

The *expected value* (or *expectation*) of a random variable is the theoretical mean of the random variable.

$$E(X) = \mu$$

To calculate the expected value of a discrete random variable X:

$$E(X) = \sum_{\text{all } x} x \cdot p(x)$$

The variance of X:

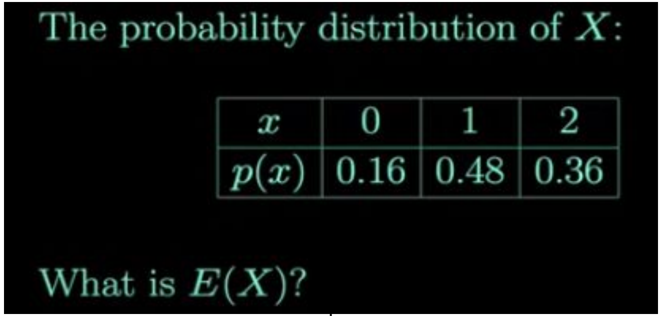
$$\sigma^2 = E[(X-\mu)^2] = \sum_{\text{all } x} (x-\mu)^2 \cdot p(x)$$

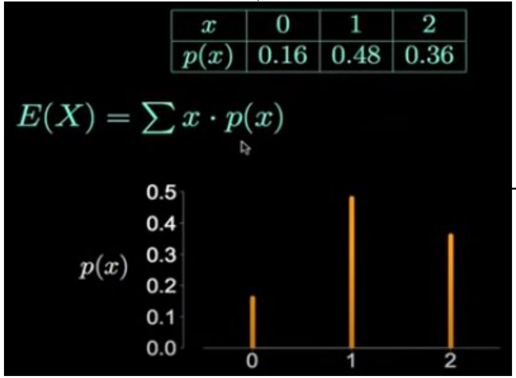
The expectation of
the squared distance of
X from its mean

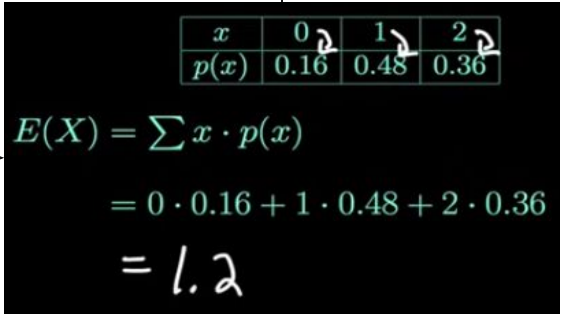
$$E[(X-\mu)^2] = E(X^2) - [E(X)]^2$$

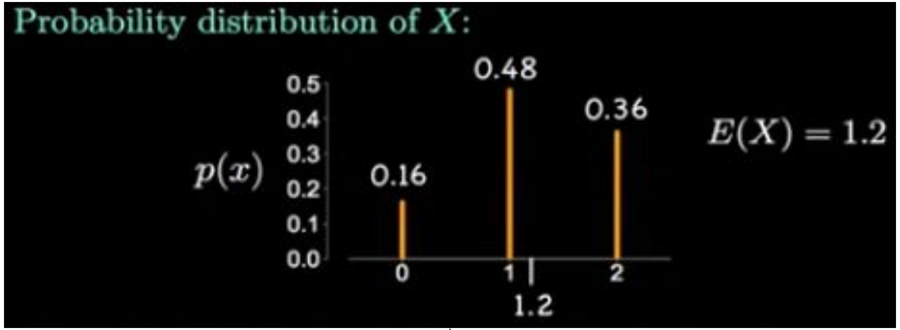
$$= E(X^2) - \mu^2$$

Misal anda memiliki koin dengan probabilitas nilai ‘kepala’ 0.6. Berapa nilai Ekspektasi dan varians jika dilempar 2 kali?









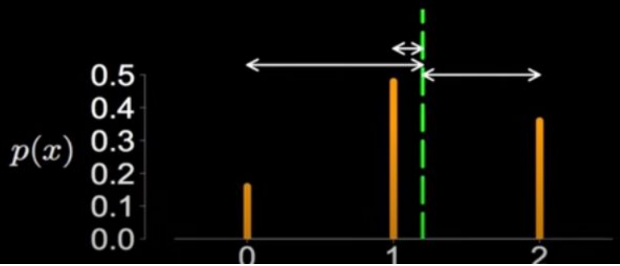
x	0^2	1^2	2^2
$p(x)$	0.16	0.48	0.36

$$E(X^2) = \sum x^2 \cdot p(x)$$
$$= 0^2 \cdot 0.16 + 1^2 \cdot 0.48 + 2^2 \cdot 0.36$$
$$= 1.92$$

What is the variance of X?

$$\sigma^2 = E[(X-\mu)^2]$$

The expectation of the squared distanc
from the mean



$(0-1.2)^2 \quad (1-1.2)^2 \quad (2-1.2)^2$

$\mu = 1.2$

x	0	1	2
$p(x)$	0.16	0.48	0.36

$$E[(X-\mu)^2] = \sum (x-\mu)^2 \cdot p(x)$$
$$= (0-1.2)^2 \cdot 0.16$$
$$+(1-1.2)^2 \cdot 0.48$$
$$+(2-1.2)^2 \cdot 0.36$$
$$= 0.48$$

$\sigma^2 = 0.48$
$\sigma = \sqrt{0.48}$

x	0	1	2
$p(x)$	0.16	0.48	0.36

$E(X) = 1.2$
$E(X^2) = 1.92$
$E[(X-\mu)^2] = 0.48$

$$E[(X-\mu)^2] = E(X^2) - [E(X)]^2 \checkmark$$

$$1.92 - 1.2^2 = 0.48$$

Tugas#2

1. Cari tahu lebih lanjut apa itu Kullback Leibler (KL) Divergence. Apa hubungan KL-Divergence dengan utility function? pada kasus apa saja kita dapat menggunakan KL-Divergence sebagai utility function?
2. Selain utility function yang telah disebutkan, sebutkan dan jelaskan utility function lainnya?
3. Sebutkan algoritma-algoritma machine learning yang dapat mengaproksimasi gaussian mixture model? Apa yang begitu spesial dari GMM sehingga machine learning mencoba mengaproksimasi GMM?