Foundations of Machine Learning
AI2000 and AI5000

FoML-03
Probability - Expectation, Variance and Gaussian Distribution

Dr. Konda Reddy Mopuri
Department of AI, IIT Hyderabad
July-Nov 2025

భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

DiL
Data-driven Intelligence
& Learning Lab

So far in FoML

- What is ML and the learning paradigms
- Probability refresher
    - Sum rule, product rule, Random variables, Bayes Theorem, Independence

Expectation, Variance and the Gaussian Distribution

భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

DiL
Data-driven Intelligence
& Learning Lab

Expectation

- Random variable X and a function $f: X \rightarrow \mathbb{R}$

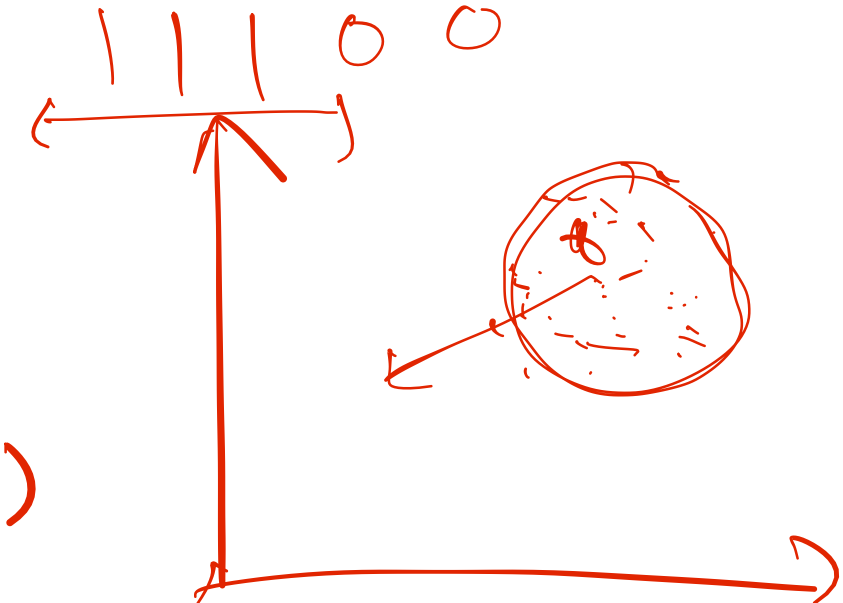
$$\mathbb{E}[f] = \mathbb{E}_{x \sim p(X)}[f(x)] = \int_X f(x)\, p(x)\, dx$$

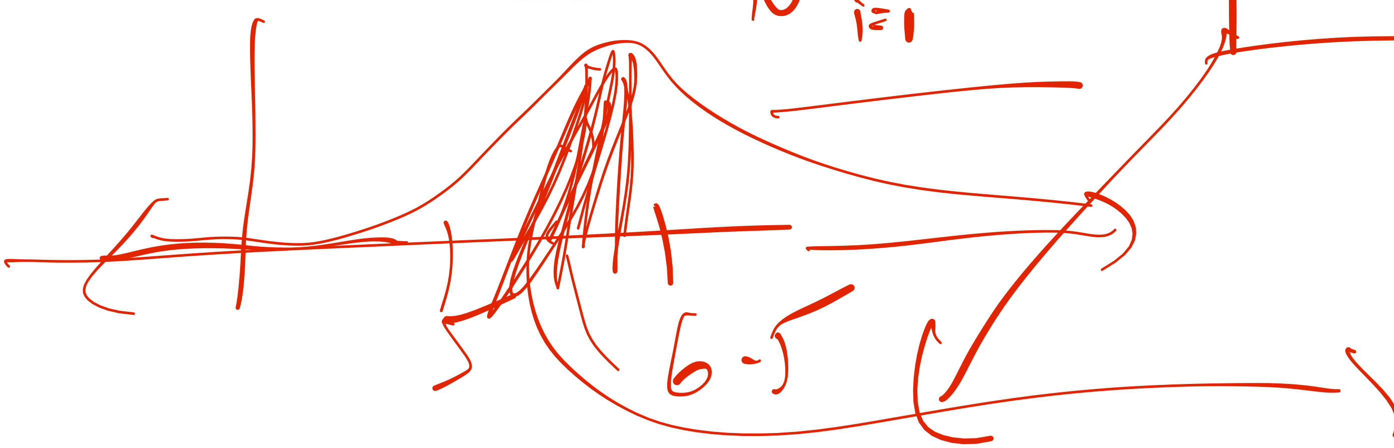
$$\sum_x f(x)\, p(x)$$

Expectation

- For N points drawn from p(X)

$$\mathbb{E}[f] = \frac{1}{N}\sum_{i=1}^{N} f(x_i)$$





Expectation

- Conditional expectation

$$\mathbb{E}[f/y] = \mathbb{E}_{x\sim p(X/Y=y)}[f(x)] \quad = \sum_{a \in A} f(a)\, P(X=a \mid Y=y)$$

$$f(\alpha X + \beta Y) = \alpha f(X) + \beta f(Y)$$

భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Variance

- Expected quadratic distance between f and its mean $\mathbb{E}[f]$

$$\mathrm{var}(f) = \mathbb{E}_{x \sim p(x)} \left(f(x) - \mathbb{E}[f(x)] \right)^2$$

$$\left(x - \mathbb{E}[x] \right)^2$$

$$= \mathbb{E}[f(x)^2] - \mathbb{E}[f(x)]^2$$

Covariance

- Measures the extent to which two random variables X and Y vary together

$$\text{cov}[X,Y] = \mathbb{E}\left[(X - E(X))(Y - E(Y))\right]$$

$$= \underbrace{\mathbb{E}[XY]}_{P(X,Y)} - \underbrace{\mathbb{E}(X)}_{P(X)}\ \underbrace{E[Y]}_{P(Y)}$$

$P(x,y)$

Covariance

$D \times 1$

- X and Y are vectors of random variables
- Covariance matrix

$$\mathrm{cov}[X,Y] = E_{P(X,Y)}\left[X - E(X)\right]\left[Y - E(Y)\right]^T$$

(i, j)

$D \times D$

Covariance

- Between independent variables

$$\text{cov}[X,Y] = E[xy] - E[x]E[y]$$

$$= E(x)E(y) -$$

$$= 0$$

Gaussian Distribution

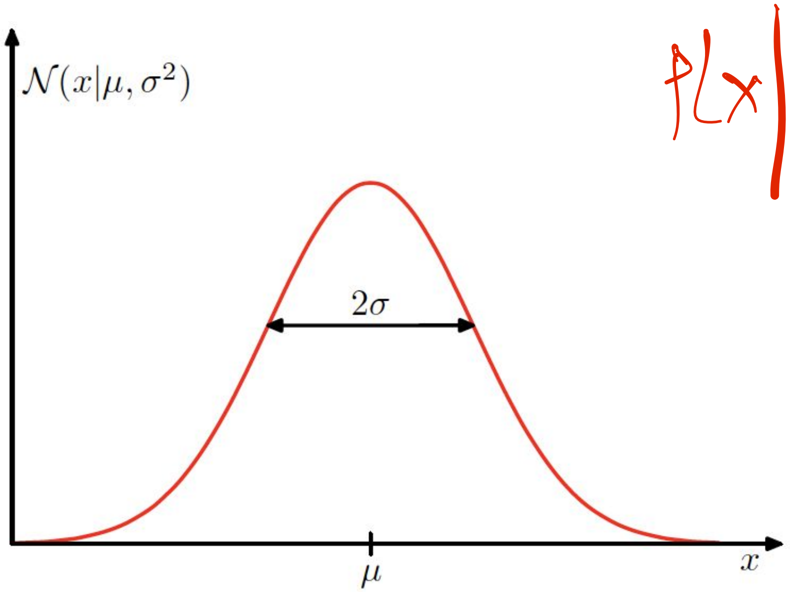
భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

DiL
Data-driven Intelligence
& Learning Lab

Gaussian Distribution

$\mathbb{E}_{x \sim \mathcal{N}(\mu, \sigma^2)}$ $\rightarrow \mu$

$\rightarrow \sigma^2$



$$p(x|\mu, \sigma^2) \quad \mathcal{N}(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

$$x \sim \mathcal{N}(x|\mu, \sigma^2)$$

$$\mathbb{E}[x] = \mu \qquad \mathrm{Var}(x) = \sigma^2$$

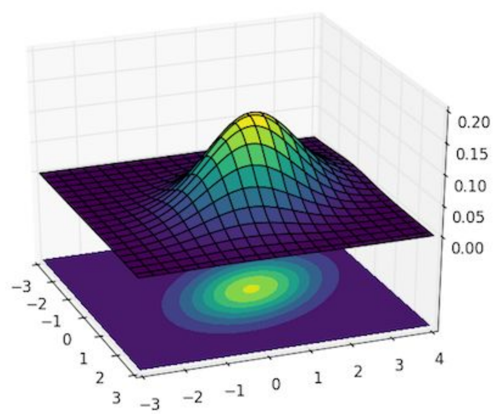
భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

DiL
Data-driven Intelligence & Learning Lab

Multivariate Gaussian Distribution

- D-dimensional vector $\mathbf{x} = (x_1, x_2, \dots x_D)^T$

$$\mathcal{N}(\mathbf{x}|\mu, \Sigma) = \frac{1}{(2\pi)^{D/2}|\mathbf{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$$



Next

Maximum Likelihood Principle