→ x (lower case bold letter) : A vector

→ All vectors are assumed to be col vectors

→ Upper case bold Roman letters eg: X denote matrices

→ $(w_1, w_2, \ldots, w_M)$: A row vector with 'm' elts

→ Conventionally col. vectors: $w = (w_1 \ldots w_M)^T$

$[a, b]$: closed interval from a to b inclusive of 'a' & 'b'.

(a, b) : open interval; excluding 'a' & 'b'.

$[a, b)$: 'a' inclusive & 'b' excluded.

I_M : $(M \times M)$ identity matrix.

$E_x[f(x, y)]$: Expectation of a func $f(x, y)$ w.r.t. to a R.V. 'x'

Conditional Expectation: $E_x[f(x) | z]$

It denotes of 'x' is conditioned on 'z'

Variance : $\text{var}[f(x)]$

Covariance : $\text{cov}[x, y]$

$\text{cov}[x, x] = \text{cov}[x, x]$

$x = (x_1, x_2, \cdots, x_D)^T$:

N values $(x_1, x_2, \cdots, x_N)$ of

a D-dimensional vector

$x = (x_1, \cdots, x_D)^T$

$X = (x_1, \cdots, x_N)^T$: A training set

consisting of N observations of x.

$$y(x, w) = w_0 + w_1 x + w_2 x^2 + \cdots + w_M x^M$$
$$= \sum_{j=0}^{M} w_j \cdot x^j$$

↓

A polynomial function of M^{th} order

w = vector of polynomial coeff's $w_0, w_1, \ldots w_M$.

Coeff. values will be determined by fitting the polynomial to the training data.

This can be done by minimizing an "Error fnc" that measures the misfit b/w the fnc $y(x, w)$, for any given value of w & the training set data points.

$$E(w) = \frac{1}{2} \sum_{n=1}^{N} \{y(x_n, w) - t_n\}^2$$

@ Regularization:

→ A technique to control the over fitting phenomenon

→ It adds a penalty ~~fnc~~ term to the err. fnc() in order to discourage the coeff.s from reaching

larger values.

$$\tilde{E}(w) = \frac{1}{2}\sum_{n=1}^{N}\{y(x_n, w) - t_n\}^2 + \frac{\lambda}{2}\|w\|^2$$

where $\|w\|^2 = w^T \cdot w$

$$= w_0^2 + w_1^2 + \cdots + w_M^2$$

The coeff. λ governs the relative importance of the regularization term compared with the SSE term.

Joint probability: The prob. that 'x' will take the value x_i & 'y' will take the value y_j is written as

$$P(X = x_i, Y = y_j)$$

and is called the "Joint prob."

of $X = x_i$ & $Y = y_j$.

Given by the no. of points falling in the cell (i,j) as a fraction of the total no. of points:

$$p(X = x_i, Y = y_j) = \frac{n_{ij}}{N}$$

VVIMP: X, Y : 2 R.V.'s

→ X takes two values $\{x_i\}$; $i = 1, 2, \cdots, M$

→ Y takes two values $\{y_j\}$; $j = 1, 2, \cdots, L$

→ n_{ij}: n_{ij} denotes the no. of points in the corresponding cell of the array.

$i \rightarrow$ column; j: row (n_{ij})
(c_i)

(6) Relationship b/w Conditional Prob.'s

$$p(y|x) = \frac{p(x|y) \cdot p(y)}{p(x)}$$

(Bayes theorem)

Dr in Bayes theorem is a

→ Normalization Constant

→ Reqd to ensure that the sum of the conditional prob. on LHS over all values of y equals 'One'.

P.T.O.

⑥ Gaussian Distrib: one of the most imp prob distrib

→ Used for continuous var.s

→ Also called "Normal Distrib."

→ Let x = a single real valued var.

Formula

$$\mathcal{N}(x|\mu,\sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}} \cdot \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

Governed by 2 parameters:

→ μ = Mean

→ σ^2 = variance

→ $\sqrt{\sigma^2} = \sigma$ = Std. deviation

→ $\beta = \frac{1}{\sigma^2}$ = Precision

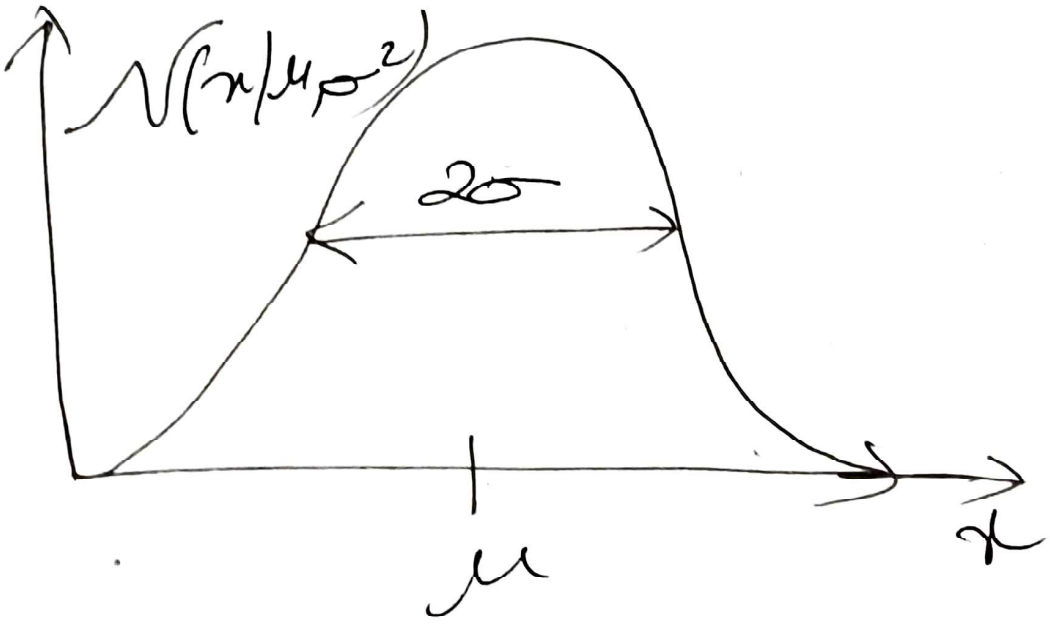
The Gaussian distribution satisfies:

$$\mathcal{N}(x|\mu,\sigma^2) > 0$$

It is straightforward to show that the Gaussian is normalized. so that

$$\int_{-\infty}^{\infty} \mathcal{N}(x|\mu,\sigma^2)\,dx = 1$$

Plot of univariate Gaussian



we can readily find expectations of funcs of 'x' under (1).

In particular, the avg. value of 'x' is given by:

$$\mathbb{E}[x] = \int_{-\infty}^{\infty} \mathcal{N}(x|\mu,\sigma^2)\, x\, dx$$

$$= \mu$$

μ = Represents the avg. value of 'x' under the distrib., it is referred to as the mean.

Hlly for the 2nd order moment:

$$\mathbb{E}[x^2] = \int_{-\infty}^{\infty} \mathcal{N}(x|\mu,\sigma^2)\, x^2\, dx$$

$$= \mu^2 + \sigma^2$$

Variance of x:

$$\text{var}[x] = E[x^2] - E[x]^2$$

$$= (\mu^2 + \sigma^2) - (\mu)^2$$

$$= \sigma^2$$

$\therefore$ Variance Parameter

The GD defined over a D-dimensional vector 'x' of continuous variables is given by:

$$N(x|\mu, \Sigma) \quad \cancel{\frac{1}{(2\pi D)^{1/2}}}$$

$$= \frac{1}{(2\pi)^{D/2}} \cdot \frac{1}{|\Sigma|^{1/2}} \cdot$$

$$\exp\left\{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right\}$$

μ = a D-dimensional vector called the mean

$\Sigma = D \times D$ matrix called covariance

$|\Sigma|$ = Determinant of Σ

we have a data set of observations:

$$X = (x_1, \ldots, x_N)^T$$

↓

Represents N observations of the scalar variable 'x'.

i.e: 'X' is a matrix of the form:

	f_1	...	f_D
x_1			
x_2			
⋮			
x_N			

1D - Dimensional features

i.e. $x_1, \ldots, x_N$ are no. of observations of x

Assumptions: The observations are drawn independently from a G.D. where 'μ' & 'σ^2' are continuous;

& we would like to determine these parameters from the data set.

i.i.d: data points that are drawn independently from the same distrib.

→ (01) Unconditional prob:

Marginal prob: The prob. of an event irrespective of the outcome of another variable.

eg: A coined dragon is red ($p(\text{red}) = \frac{2}{4} = 0.5$) $\quad p(A)$

Conditional prob: The prob. of one event occuring in the presence of a second event.

Joint prob: prob. of 2 events occuring simultaneously.

P(A): prob. of event A (or) P(x=A)
↓
a R.V.

P(x): 'x' is a R.V;
P(x) assigns a prob. to all values of x.
↓
(a fn(c))

$$\text{Probability} = \frac{\text{No. of desired outcomes}}{\text{Total no. of possible outcomes}}$$

Note: Sum of prob.s of all outcomes should be equal to one.

{Prob. of a certain outcome = 1.0}

Complement: prob. of an event not occurring.

Denoted as $\boxed{1 - P(A)}$

(A) (1) The prob. of a row of data is the Joint prob. across each I/P var.

(2) prob. of a specific value of one I/P. var. is the marginal prob. across the values of other I/P vars.

(3) The "predictive model" itself is an estimate of the condn. prob. of an O/P given an I/P eg.

Joint prob. Distrib: The joint prob. of 2 or more R.V's.

eg: P(A and B)
= P(A given B) * P(B)

(6) Fundamental Rule of prob.

(OR)

Product Rule (OR) Chain Rule

= Calculation of Joint prob.

Joint prob is symmetrical i.e.

$$P(A \text{ and } B) = P(B \text{ and } A)$$

$$P(A \text{ and } B) = P(A \text{ given } B) * P(B)$$

$$= P(B \text{ given } A) * P(A)$$

$$P(A \text{ given } B) = \frac{P(A \text{ and } B)}{P(B)}$$

Joint prob: $P(A \text{ and } B) = P(A) * P(B)$

Note: The "marginal prob." for an event for an independent R.V. is simply the prob of the event.

Cond. prob: $P(A \text{ given } B) = P(A)$

$\Rightarrow$ prob. of B has no effect on 'A'.