# Probability & Statistics

→ Basic Probability review (Done in MTL106)

## Multivariate Gaussian distribution ★ very imp.

$x \in \mathbb{R}^n$ (vector) ; $\mu \in \mathbb{R}^n$ (mean) ; $\Sigma \in \mathbb{R}^{n \times n}$ (covariance matrix)

$$\boxed{p(x; \mu, \Sigma) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right)}$$

$\Sigma \to$ full rank & positive S.D

## Conditional Expectation

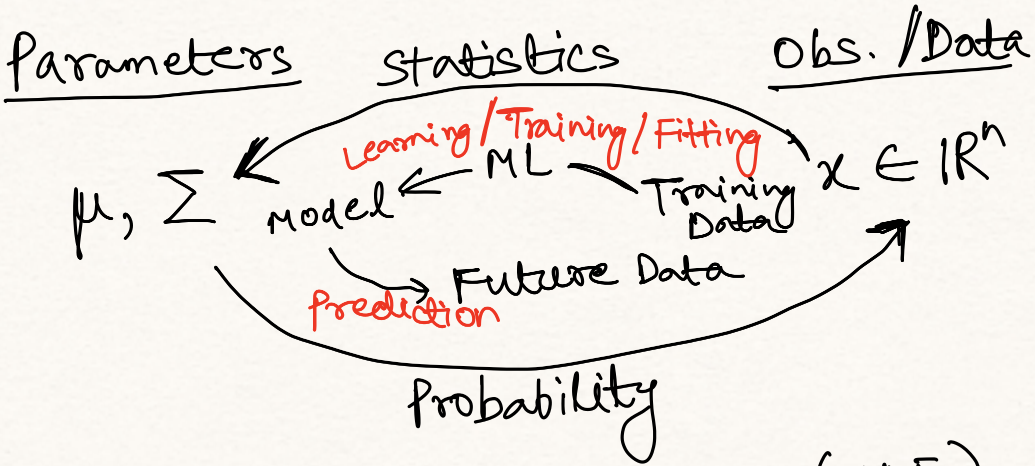
$E[X|Y] \to$ r.v

$E[X|Y=y] \to$ func$^n$ of $y$.

★ $E[X] = E[E[X|Y]]$

$$p(a|b) = \frac{p(b|a) \cdot p(a)}{p(b)}$$

$$p(a|b,c) = \frac{p(b|a,c) \cdot p(a|c)}{p(b|c)}$$




- Max$^m$ Likelihood Estimation. (MLE)

**MLE** (Gaussian)

$x \in \mathbb{R}^d$

$X = \{x^{(1)}, \ldots, x^{(n)}\}$ sampled independently & identically distributed (i.i.d)

$x^{(i)}$ – $i$th example $\in \mathbb{R}^d$

Probability Density $\quad p(x; \mu, \Sigma) = \dfrac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right]$

$$p(x^{(1)}, \ldots, x^{(n)}; \mu, \Sigma) = \prod_{i=1}^{n} \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(x^{(i)} - \mu)^T \Sigma^{-1} (x^{(i)} - \mu)\right]$$

n i.i.d vectors

Prob. of data given parameters

$\left[p(x^{(1)}, \ldots, x^{(n)}) = \prod_{i=1}^{n} p(x^{(i)}) \text{ for i.i.d}\right]$

Likelihood func$^n$:

$$L(\mu, \Sigma; x^{(1)}, \ldots, x^{(n)}) = \boxed{\prod_{i=1}^{n} \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(x^{(i)} - \mu)^T \Sigma^{-1} (x^{(i)} - \mu)\right]}$$

Likelihood of parameters given data

★ Vol. under likelihood func$^n$ may not be 1.

$$L(\theta; X) = \prod_{i=1}^{n} L(\theta; x^{(i)})$$

$$\hat{\theta}_{MLE} = \arg\max_{\theta} \prod_{i=1}^{n} L(\theta; x^{(i)})$$

$$= \arg\max_{\theta} \log \prod_{i=1}^{n} L(\theta; x^{(i)})$$

$$= \arg\max_{\theta} \sum_{i=1}^{n} \ell(\theta; x^{(i)})$$

$$\ell(\theta) = \log L(\theta)$$

---

Gaussian

$$\hat{\mu}, \hat{\Sigma} = \arg\max_{\mu, \Sigma} \sum_{i=1}^{n} \log\left[\frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(x^{(i)}-\mu)^T \Sigma^{-1} (x^{(i)}-\mu)\right]\right]$$

$$= \arg\max_{\mu, \Sigma} \left[\sum_{i=1}^{n} K - \frac{1}{2}\log|\Sigma| - \frac{1}{2}(x^{(i)}-\mu)^T \Sigma^{-1} (x^{(i)}-\mu)\right]$$

Now we have to diff. to find max.

| $\mu$ | $\Sigma$ |
|---|---|

$$\nabla_\mu \sum_{i=1}^{n} \left[ K - \frac{1}{2}\log|\Sigma| - \frac{1}{2}(x^{(i)}-\mu)^T \Sigma^{-1}(x^{(i)}-\mu) \right]$$

$$\nabla_\mu \sum_{i=1}^{n} -\frac{1}{2} \left[ x^{(i)T}\Sigma^{-1}x^{(i)} - x^{(i)T}\Sigma^{-1}\mu - \mu^T\Sigma^{-1}x^{(i)} + \mu^T\Sigma^{-1}\mu \right]$$

$$\left[ \nabla_x x^T A x = 2Ax \text{ if } A \text{ symm.} \quad \Sigma \text{ is symm.} \right]$$

$$\nabla_\mu \sum_{i=1}^{n} -\frac{1}{2} \left[ -2(\Sigma^{-1}x^{(i)})^T \mu + \mu^T \Sigma^{-1} \mu \right]$$

$$\sum_{i=1}^{n} \left( \Sigma^{-1}x^{(i)} - \Sigma^{-1}\mu \right) = 0$$

$$n\Sigma^{-1}\mu = \sum_{i=1}^{n} \Sigma^{-1}x^{(i)}$$

$$\Sigma^{-1}\mu = \Sigma^{-1}\left[ \frac{1}{n}\sum_{i=1}^{n} x^{(i)} \right]$$

$$\boxed{\mu = \frac{1}{n}\sum_{i=1}^{n} x^{(i)}}$$

$$S = \Sigma^{-1}$$

$$\nabla_S \ell = 0 \Leftrightarrow \nabla_{\Sigma^{-1}} \ell = 0$$

$$\nabla_S \sum_{i=1}^{n} \frac{1}{2}\log|S| - \frac{1}{2}(x^{(i)}-\mu)^T S (x^{(i)}-\mu)$$

$$\frac{1}{2}\left[ nS^{-1} - \sum_{i=1}^{n}(x^{(i)}-\mu)(x^{(i)}-\mu)^T \right] = 0$$

$$\left[ \nabla_A x^T A x = xx^T \right]$$

$$S^{-1} = \frac{1}{n}\sum_{i=1}^{n}(x^{(i)}-\mu)(x^{(i)}-\mu)^T$$

$$\boxed{\Sigma = \frac{1}{n}\sum_{i=1}^{n}(x^{(i)}-\mu)(x^{(i)}-\mu)^T}$$