Day 05

Univariate & Multivariate Gaussian Random Variables

Readings: Bishop PRML Sec 1.2.4
"The Gaussian"
Basic PDF, ML estimator

Bishop PRML Sec 2.3
"The Gaussian"
Deeper concepts for multivariate Gaussian

Goals:

day 5

Univariate Gaussian
- PDF function
- mean and variance
- ML estimators

coming soon
day 6

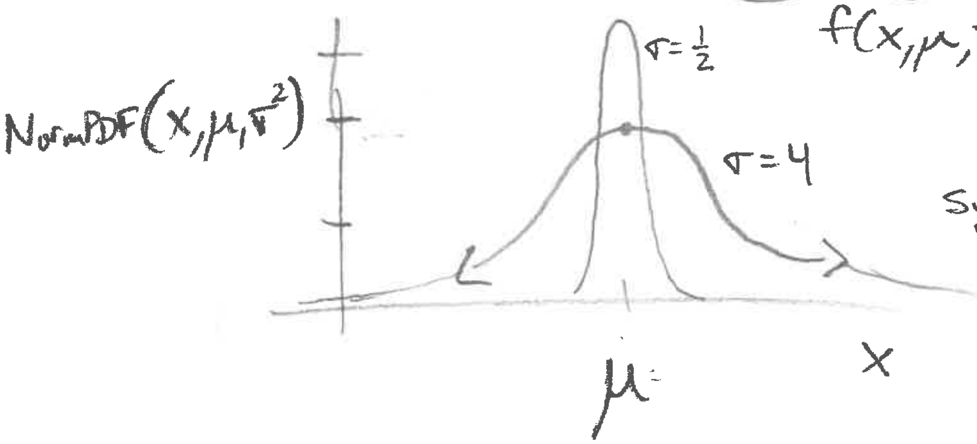
Multivariate Gaussian
- Understanding covariance matrices
- PDF function
- ML estimator

Gaussian - Gaussian models
- Posterior Evidence MAP

Graphical interpretation of Normal PDF

$$\text{NormPDF}(x \mid \mu, \sigma^2) \propto \underbrace{e^{-\frac{1}{2}\frac{1}{\sigma^2}(x-\mu)^2}}_{f(x,\mu,\sigma^2)}$$

e^{-a^2} is always ≤ 1



symmetric function around $x = \mu$

maximum at $x^* = \mu$

σ controls width ("scale")

Compute variance of X?

Recall: $\text{Var}[X] = \underbrace{\mathbb{E}[X^2]}_{\text{(a) need this}} - \underbrace{\mathbb{E}[X]}_{\text{(b) we already know this is } \mu}{}^2$

(for any random variable)

$$\mathbb{E}[X^2] = \int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} dx$$

$$= \mu^2 + \sigma^2$$ (after similar math derivation to $\mathbb{E}[X] = \mu$)

hus,

$$\text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

$$= \mu^2 + \sigma^2 - (\mu)^2 = \boxed{\sigma^2}$$

Univariate Gaussian distribution

Let X be a univariate Gaussian r.v.
("Gaussian" also called "Normal")
then we have

sample space : whole real line ($\mathbb{R}^1$)

parameters : $\mu \in \mathbb{R}$ called "mean" or "location"
$\sigma^2 > 0$ called "variance" or "scale"

P.D.F. is :

$$\mathrm{NormPDF}(x \mid \mu, \sigma^2) = \underbrace{\frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\sigma^2}}}_{\text{const wrt } x} \underbrace{e^{-\frac{1}{2}\frac{1}{\sigma^2}(x-\mu)^2}}_{f(x,\mu,\sigma^2)}$$

Note identity: (*) $\int e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} dx = \sigma\sqrt{2\pi}$

Expectation of X :

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{1}{\sigma^2}(x-\mu)^2} dx$$

Change of variables

$t = \frac{x-\mu}{\sigma} \curvearrowright x = \sigma t + \mu$

$dt = \frac{1}{\sigma} dx$

$$= \int_{-\infty}^{\infty} (\sigma t + \mu) \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}t^2} \cdot \sigma dt$$

$f(t) = te^{-\frac{1}{2}t^2}$

$f(-t) = -te^{-\frac{1}{2}(-t)^2}$

$= -f(t)$

odd function

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \left[\sigma^2 \int_{-\infty}^{\infty} te^{-\frac{1}{2}t^2} dt + \mu\sigma \int_{-\infty}^{\infty} e^{-\frac{1}{2}t^2} dt \right]$$

use the * identity

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \left[\sigma^2 \cdot \underset{\text{integral of odd function}}{\cancel{0}} \right] + \cancel{\sqrt{2\pi}} \cdot \cancel{\frac{1}{\sigma}} \cdot \mu \cdot \cancel{\sigma} \cdot \cancel{\sqrt{2\pi}} = \mu$$

Consider observing N data measurements $x_1, \cdots x_N$, where each $x_n \in \mathbb{R}$

If we assume these are indep. and identically distribute[d]

$$p(x_1 \cdots x_N) = \prod_{n=1}^{N} p(x_n | \mu, \sigma^2) = \prod_{n=1}^{N} \text{NormPDF}(x_n | \mu, \sigma^2)$$

we can try to estimate the (unknown) parameters μ and σ^2 by maximizig likelihood

$$\mu_{ML}, \sigma_{ML} = \underset{\substack{\mu \in \mathbb{R} \\ \sigma^2 > 0}}{\operatorname{argmax}} \sum_{n=1}^{N} \log \text{NormPDF}(x_n | \mu, \sigma^2)$$

$$= \operatorname{argmax} \sum_{n=1}^{N} \log \left[\frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \exp\left[-\frac{1}{2} \frac{1}{\sigma^2} (x - \mu)^2 \right] \right]$$

$$= \operatorname{argmax} \underbrace{- \frac{N}{2} \log [2\pi] - N \log \sigma - \frac{1}{2} \frac{1}{\sigma^2} \overbrace{\sum_{n=1}^{N} (x_n - \mu)^2}^{\text{sum of squared distances}}}_{\text{call this } \mathcal{L}(\mu, \sigma)}$$

[T]o solve, we find local maxima via std. calculus methods
Set first deriv. wrt μ, σ to $\varnothing$, then solve.

Sensibly, we set μ to the empirical mean of t[he] dat[a]

$$\frac{\partial}{\partial \mu} \mathcal{L}(\mu, \sigma) = \frac{\partial}{\partial \mu} \left[-\frac{1}{2} \frac{1}{\sigma^2} \sum_n (x_n^2 - 2\mu x_n + \mu^2) \right]$$

$$= + \frac{1}{\sigma^2} \sum_{n=1}^{N} x_n - \frac{1}{\cancel{2}} \frac{1}{\sigma^2} \cancel{2} N \mu = 0$$

$$\sum x_n = N\mu \longrightarrow \boxed{\mu^* = \frac{\sum_n x_n}{N}}$$

Similarly, solve for σ:

$$\frac{\partial}{\partial \sigma} \mathcal{L}(\mu, \sigma) = 0$$

sum of squared differences, a function of both $x_1 \cdots x_N$ and mean μ

$$\frac{\partial}{\partial \sigma}\left[-N \log \sigma - \frac{1}{2}\frac{1}{\sigma^2} SSD(x,\mu) \right] = 0$$

$$-\frac{N}{\sigma} - \frac{1}{2} SSD(x,\mu) \frac{\partial}{\partial \sigma}\left[\sigma^{-2}\right] = 0$$

$$-\frac{N}{\sigma} - \cancel{\tfrac{1}{2}} SSD \cdot (\cancel{-2}) \sigma^{-3} = 0$$

$$-N\sigma^{-1} + SSD\,\sigma^{-3} = 0$$

$$-N\sigma^2 + SSD = 0 \rightarrow \sigma^2 = \frac{SSD(x,\mu)}{N}$$

Thus, we have:

$$\boxed{\mu_{ML} = \frac{\sum_n x_n}{N} \quad \text{and} \quad \sigma^2_{ML} = \frac{\sum_n (x_n - \mu_{ML})^2}{N}}$$

Question:

μ_{ML} and σ_{ML} are estimators given a finite sample of N examples.
If the data comes from an iid Gaussian model w/ parameters μ, σ^2
will the expected value of μ_{ML} be equal to μ? Yes
σ^2_{ML} be equal to σ^2? No!

$$\mathbb{E}_{x_n \overset{iid}{\sim} N(\mu,\sigma^2)}\left[\mu_{ML}(x_1 \cdots x_N)\right] = \mu? \quad \underline{\text{Yes}}.$$

$$\mathbb{E}_{x_n \overset{iid}{\sim} N(\mu,\sigma^2)}\left[\frac{1}{N}(x_1 + \cdots x_N)\right] = \frac{1}{N}\left(\mathbb{E}[x_1] + \cdots \mathbb{E}[x_N]\right) = \frac{N}{N}\mu = \boxed{\mu}$$

derivation that σ_{ML} is biased estimator for finite N

$$\mathbb{E}\left[\sigma^2_{ML}(x_1 \cdots x_N)\right] = \mathbb{E}\left[\frac{1}{N}\sum_{n=1}^{N}\left(x_n^2 - 2\mu_{ML}x_n + \mu_{ML}^2\right)\right]$$

$x_n \overset{iid}{\sim} N(\mu, \sigma^2)$

$$= \frac{1}{N}\sum_{n=1}^{N}\left(\mathbb{E}[x_n^2] - 2\,\mathbb{E}[\mu_{ML}(x_1 \cdots x_N)\cdot x_n] + \mathbb{E}[\mu_{ML}^2]\right)$$

$$= \frac{1}{N}\left[\sum_{n=1}^{N}\mathbb{E}[x_n^2] - 2N\,\mathbb{E}\left[\mu_{ML}(x_1 \cdots x_N)\cdot\frac{(\sum_n x_n)}{N}\right] + N\,\mathbb{E}[\mu_{ML}^2]\right]$$

$$= \frac{1}{N}\left[\sum_{n=1}^{N}\mathbb{E}[x_n^2] - 2N\,\mathbb{E}[\mu_{ML}^2] + N\,\mathbb{E}[\mu_{ML}^2]\right]$$

$$= \left[\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[x_n^2]\right] - \mathbb{E}[\mu_{ML}^2]$$

$$= \cancel{\frac{N}{N}}\mu^2 + \sigma^2 - \mathbb{E}\left[\frac{1}{N}(x_1 + x_2 + \cdots x_N)\frac{1}{N}(x_1 + \cdots x_N)\right]$$

$$- \frac{1}{N^2}\,\mathbb{E}\left[\sum_n x_n^2 + \sum_{i \neq j} x_i x_j\right]$$

$$= \mu^2 + \sigma^2 - \frac{1}{N^2}\sum_n(\mu^2 + \sigma^2) - \frac{1}{N^2}\sum_{i \neq j}\mathbb{E}[x_i x_j]$$

$$= \mu^2 + \sigma^2 - \frac{1}{N^2}\underbrace{\sum_n}_{N \text{ terms}}\mu^2 + \sigma^2 - \frac{1}{N^2}\underbrace{\sum_{i \neq j}}_{N^2 - N \text{ terms}}\mu^2$$

$$= \mu^2 - \frac{N}{N^2}\mu^2 - \frac{N^2 - N}{N^2}\mu^2 + \sigma^2 - \frac{N}{N^2}\sigma^2$$

$$= 0 \cdot \mu + \frac{N^2 - N}{N^2}\sigma^2 = \frac{N-1}{N}\sigma^2$$

Assuming $x_1 \cdots x_N \sim \mathcal{N}(\mu, \sigma^2)$

Punchline: μ_{ML} is an unbiased estimator of the mean for any finite sample:

$$\mathbb{E}[\mu_{ML}(x_1 \cdots x_N)] = \mu$$

σ_{ML} is a biased estimator of the variance for any finite sample

$$\mathbb{E}[\sigma_{ML}^2(x_1 \cdots x_N)] = \frac{N-1}{N}\sigma^2$$

When N is small, this might matter a lot

As N gets larger, $\frac{N-1}{N} \rightarrow 1.0$ and bias gradually disappears

Another problem w/ ML estimators:

If N too small, variance will be nonsensical

e.g. at $N=1$, $\sigma_{ML}^2 = \frac{(x_1 - \mu_{ML})^2}{N} = \frac{(x_1 - x_1)^2}{N} = \emptyset$ zero!