Gaussian Distribution

The Gaussian distribution is one of the most important probability distributions. It is also refered to as the Normal distribution.

Definition. Given a *mean* μ, a *standard deviation* σ whose square is called the *variance*, the *Gaussian distribution* is given by

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

We also call the $\beta = \frac{1}{\sigma^2}$ the *precision*.

Properties of Gaussian distribution.

$$\begin{aligned} \mathbb{E}[x] &= \mu \\ \mathbb{E}[x^2] &= \mu^2 + \sigma^2 \\ \operatorname{var}[x] &= \sigma^2. \end{aligned}$$

Definition. The maximum of a Gaussian distribution is called the *mode.*

Definition. Let $\boldsymbol{x}, \boldsymbol{\mu} \in \mathbb{R}^D$ and $\boldsymbol{\Sigma}$ be an invertible $D \times D$ matrix. Then, the D-dimensional Gaussian distribution is given by

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

Definition. Data points drawn independently from the same distribution is called *independent and identically distributed*, abbreviated to *i.i.d.*

Drawing N observations $\boldsymbol{x} = \{x_1, ..., x_N\}$ from a Gaussian distribution of unknown mean and standard deviation is an i.i.d. Since we have the property $P(X,Y) = P(X)P(Y)$ for independent variables, it follows that for a given μ and σ,

$$p(\boldsymbol{x}|\mu,\sigma^2) = \prod_{n=1}^{N} \mathcal{N}(x_n|\mu,\sigma^2).$$

We can then maximise the log of the formula above and get the *sample mean* μ_{ML} as the maximal mean which turns out to just be the average of every value in $\boldsymbol{x}$. We also get the *sample variance* by maximising σ^2, and it turns out

$$\sigma^2_{\text{ML}} = \frac{1}{N-1} \sum_{n=1}^{N} (x_n - \mu_{\text{ML}}).$$

There are limitations in maximal likelihood, as the above approach systematically underestimates the variance, which is a phenomena called *bias*. We first think of μ_{ML}

and σ^2_{ML} as functions of $x_1, \ldots, x_n$. Then, taking μ and σ^2 as the actual parameters from the Gaussian distribution that generated the data points,

$$\mathbb{E}[\mu_{\mathrm{ML}}] = \mu, \mathbb{E}[\sigma^2_{\mathrm{ML}}] = \frac{N-1}{N}\sigma^2.$$

Hence the variance is underestimated with a factor, which becomes nonexistent as N approaches infinity.