Regression II

COMP9417, 22T2

1 Stats, stats, stats . . .

Stats, stats, stats . . .

Probabilty Distribution

A probability distribution represents the probability we see a value x in a sample X. We denote this as $P(X = x)$.

We typically assume our sample are i.i.d (independent and identically distributed), helping us reduce the complexity of the problem and apply statistically supported conclusions.

The tutorial discusses the Gaussian/Normal and Bernoulli distributions.

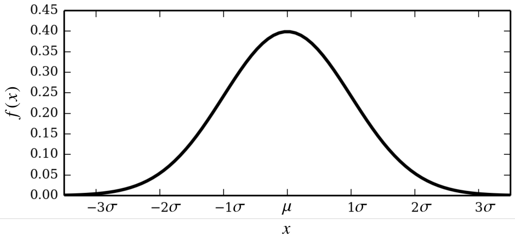
Gaussian Distribution

A standard probability distribution is the Gaussian, where:

$$\theta = (\mu, \sigma^2), \quad \mu \in \mathbb{R}, \sigma > 0$$

$$p_\theta(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

We typically write $X \sim \mathcal{N}(\mu, \sigma^2)$ to say X is normally distributed with mean μ and variance σ^2.



Maximum Likelihood Estimation

Maximum likelihood estimation is the process of estimating the parameters of a distribution of sample data by maximising the overall likelihood of the samples occuring in the distribution.

$$
\begin{aligned}
\text{Prob of observing } X_1, \cdots, X_n &= \text{Prob of observing } X_1 \times \cdots \times \text{Prob of observing } X_n \\
&= p_\theta(X_1) \times \cdots \times p_\theta(X_n) \\
&= \prod_{i=1}^{n} p_\theta(X_i) \\
&=: L(\theta) \qquad \text{this is our likelihood function}
\end{aligned}
$$

To make life easier, we typically work with the log of the likelihood function (log-likelihood). As log is a strictly increasing function, the maximisation of $L(\theta)$ and $\log(L(\theta))$ give us the same result.

$$L(\theta) = \prod_{i=1}^{n} p_\theta(X_i)$$

$$\log(L(\theta)) = \log \prod_{i=1}^{n} p_\theta(X_i)$$

$$= \sum_{i=1}^{n} \log p_\theta(X_i)$$

This makes differentiating, and therefore maximising much easier.

1a

Problem: Given $X_1, \cdots, X_n \sim N(\mu, 1)$, find $\hat{\mu}_{\text{MLE}}$.

First, we define our likelihood function:

$$
\begin{aligned}
\log L(\mu) &= \log\left(\prod_{i=1}^{n} p_\theta(X_i)\right) \\
&= \log\left(\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(X_i - \mu)^2\right)\right) \\
&= -\frac{1}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}(X_i - \mu)^2
\end{aligned}
$$

Next, we differentiate with respect to our parameter μ,

$$\frac{\partial \log L(\mu)}{\partial \mu} = \sum_{i=1}^{n}(X_i - \mu)$$
$$= \sum_{i=1}^{n} X_i - n\mu$$

$$\frac{\partial \log L(\mu)}{\partial \hat{\mu}} = 0 \text{ at the minimum. So,}$$

$$\sum_{i=1}^{n} X_i - n\hat{\mu} = 0$$
$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i$$
$$\hat{\mu} = \bar{X}$$