Probability Distributions: Continuous

Introduction to Data Science Algorithms

Jordan Boyd-Graber and Michael Paul

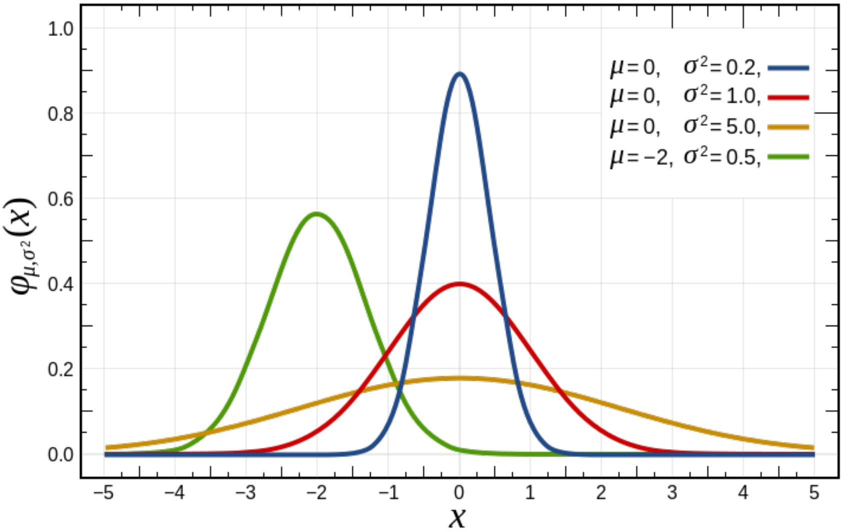
SEPTEMBER 29, 2016

The normal distribution

- The most common continuous distribution is the *normal* distribution, also called the *Gaussian* distribution.
- The density is defined by two parameters:
  - μ: the *mean* of the distribution
  - σ^2: the *variance* of the distribution (σ is the *standard deviation*)
- The normal density has a "bell curve" shape and naturally occurs in many problems.

Carl Friedrich Gauss
1777 – 1855

The normal distribution



The normal distribution

- The probability density of the normal distribution is:

$$f(x) = \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}}}_{\substack{\text{Does not}\\ \text{depend on } x}} \underbrace{\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)}_{\substack{\text{Largest when } x=\mu;\\ \text{shrinks as } x \text{ moves}\\ \text{away from } \mu}}$$

- Notation: $\exp(x) = e^x$
- If X follows a normal distribution, then $\mathbb{E}[X] = \mu$.
- The normal distribution is symmetric around μ.

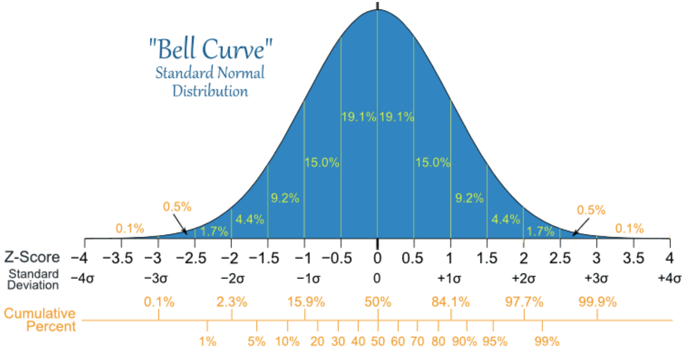
The normal distribution

- What is the probability that a value sampled from a normal distribution will be within n standard deviations from the mean?
- $P(\mu - n\sigma \leq X \leq \mu + n\sigma) = ?$

The normal distribution

- What is the probability that a value sampled from a normal distribution will be within *n* standard deviations from the mean?
- $P(\mu - n\sigma \leq X \leq \mu + n\sigma) = ?$
  $= \int_{x=\mu-n\sigma}^{\mu+n\sigma} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$
  $= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{x=\mu-n\sigma}^{\mu+n\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$

The normal distribution



Applying the normal distribution

- Most variables in the real world don't follow an exact normal distribution, but it is a very good approximation in many cases.

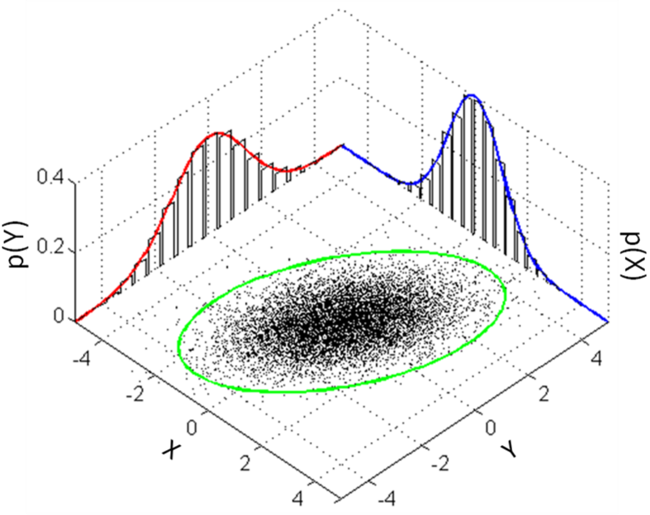
- Measurement error (e.g., from experiments) is often assumed to follow a normal distribution.
- Biological characteristics (e.g., heights of people, blood pressure measurements) tend to be normal distributed.
- Test scores

- Special case: sums of multiple random variables
  - The *central limit theorem* proves that if you take the sum of multiple randomly generated values, the sums will follow a normal distribution. (Even if the randomly generated values do not!)

Multivariate normal distribution

- What is the *joint* distribution over multiple normal variables?
- If the normal random variables are independent, the joint distribution is just the product of each individual PDF.
- But they don't have to be independent.
- We can model the joint distribution over multiple variables with the *multivariate* normal distribution.

Multivariate normal distribution



Multivariate normal distribution

- The multivariate normal distribution is a distribution over a *vector* of values $\mathbf{x}$. The mean μ is also a vector.
- In addition to the variance of each variable, each pair of variables has a *covariance*.
  - The covariance matrix for all pairs is denoted Σ.
  - The covariance indicates an association between variables. If it is positive, it means if one value increases (or decreases), the other value is also likely to increase (or decrease). If the covariance is negative, it means that if one value increases, the other is likely to decrease, and vice versa.
- $f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^k|\Sigma|}} \exp\left(-\frac{1}{2}(\mathbf{x}-\mu)^{\mathrm{T}}\Sigma^{-1}(\mathbf{x}-\mu)\right)$

Computing the Mean

- If you have observations $x_1 \dots x_N$ that come from a normal distribution, what is the mean μ?
- Formula

$$\hat{\mu} = \frac{\sum_{i=1}^{N} x_i}{N} \quad (1)$$

Computing the Variance

- If you have observations $x_1 \dots x_N$ that come from a normal distribution, what is the variance σ^2?
- Formula

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N} \tag{2}$$

Computing the Variance

- If you have observations $x_1 \dots x_N$ that come from a normal distribution, what is the variance σ^2?
- Formula

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N} \tag{2}$$

- Why? Next lecture! (Maximum likelihood)