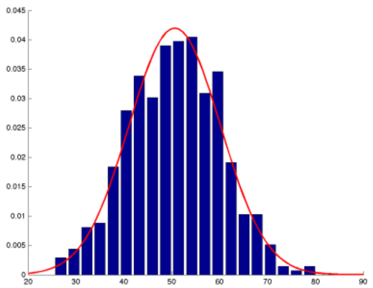
CPSC 440: Machine Learning

Multivariate Gaussian

Winter 2022

Last Time: Univariate Gaussian

- We discussed continuous density estimation with the Gaussian/normal distribution:



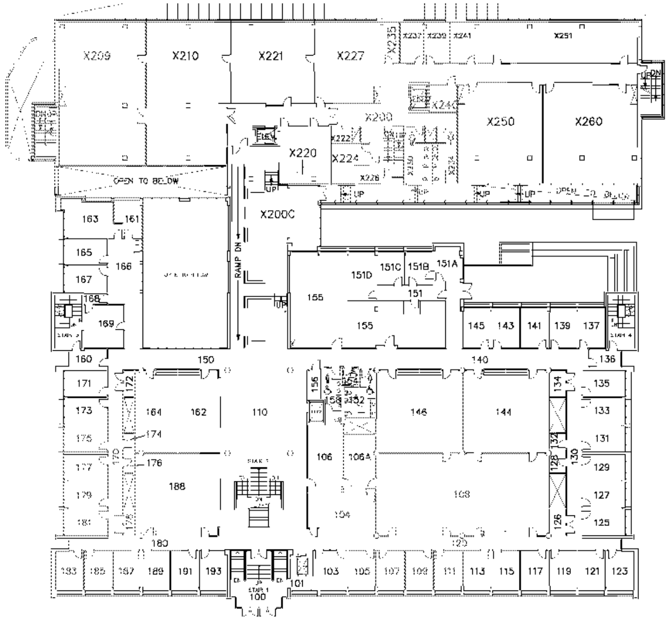
$$p(x^i \mid \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x^i-\mu)^2}{2\sigma^2}\right)$$

$$x^i \sim \mathcal{N}(\mu, \sigma^2)$$

- Parameterized in terms of mean μ and variance σ^2.
- Can convert this density to probabilities using CDF F(c).
- Can generate samples using inverse CDF: x = F^{-1}(u) ("quantile" function).
- MLE for mean is mean of data, MLE for variance of variance of data.
- For fixed variance, conjugate prior for mean is Gaussian.
  - Posterior mean converges from prior mean to mean of data as 'n' increases.
  - Posterior variance converges to 0 as 'n' increase.
  - Posterior predictive is also a Gaussian (variance does not go below σ^2).

Motivation: Modeling Air Quality

- We want to model "air quality" in different rooms in a building.
- So we measure number of pollutant molecules (PM10, CO, O3, and so on):



Rm 1	Rm 2	Rm 3	Rm 4	Rm 5	Rm 6	Rm 7	Rm 8	Rm 9
0.1	1.4	0.2	1.8	1.0	1.0	0.1	0.1	1.1
0.2	1.3	0.1	1.9	1.1	0.9	0.1	0.1	1.1
0.1	0.3	1.4	2.0	0.7	0.3	0.1	0.2	0.4
0.1	1.1	0.2	2.1	1.1	1.1	0.1	0.3	0.5
2.7	2.6	2.5	5.1	2.4	2.8	3.2	2.5	3.1
0.1	0.4	0.2	1.8	1.3	0.4	0.1	0.4	1.0
0.1	1.2	0.2	1.8	1.4	1.1	0.7	0.7	0.5

- We want to build a model of this data, to identify patterns/problems.
  - Some rooms usually bad air quality, some usually have good air quality.
  - The quality of some rooms may be correlated (rooms are adjacent or share air supply).
  - There are also temporal correlations (we will come back to temporal correlations later).