Artificial Intelligence

Using Distributions

CS 444 – Spring 2019
Dr. Kevin Molloy
Department of Computer Science
James Madison University

Why Use Distributions?

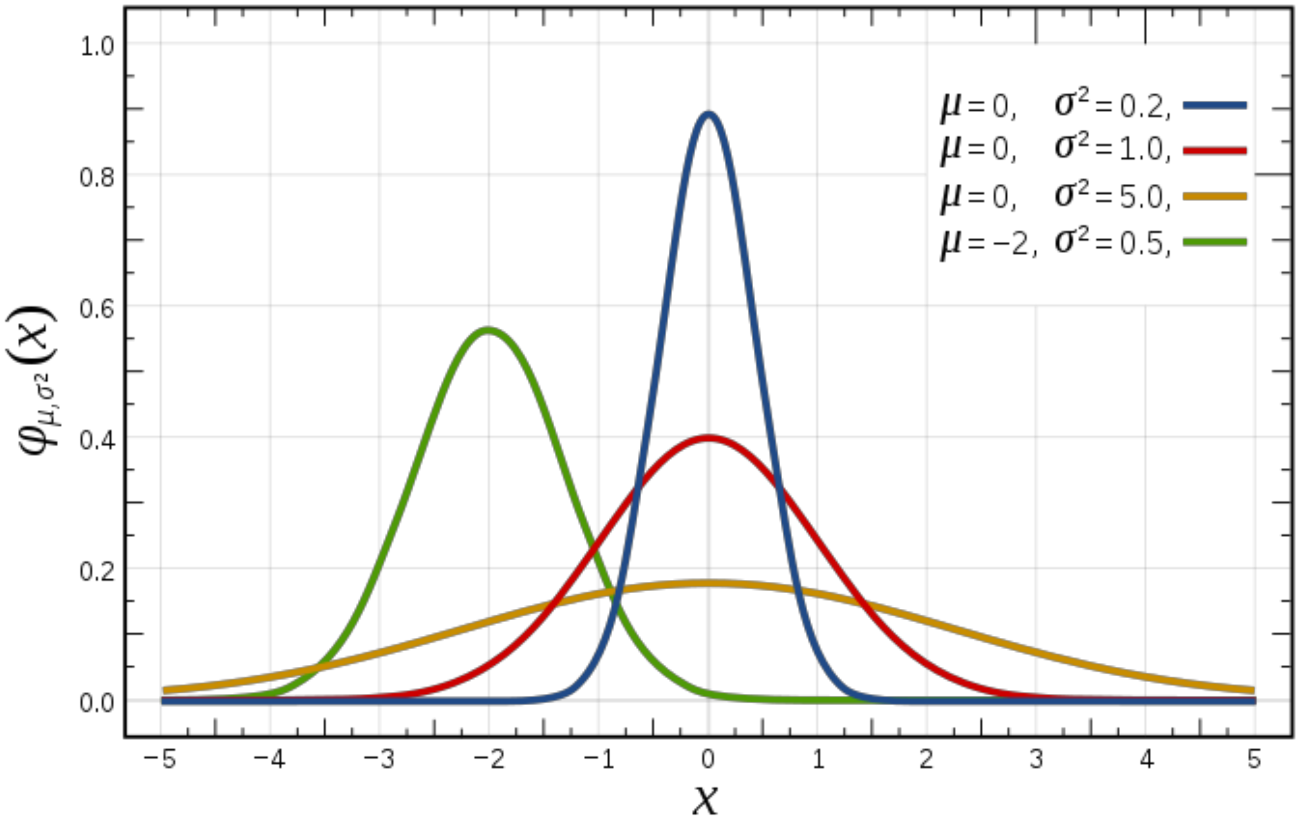
- The world is a stochastic place
- Summarizes data in a few parameters
- Provide the ability to easily sample
- Summarize belief state probabilistically

Univariate Gaussian

Gauss was a German mathematician. His work in mathematics helped locate the dwarf planet **Ceres**.

Key idea is probability theory is the central limit theorem (CLT). Briefly, it states that when independent random variables are added, their properly normalized sum tends towards a normal/gaussian distribution, even if the original variables are not normally distributed.

This, in theory, allows Gaussians to be used to applied to many problems.



Univariate Form

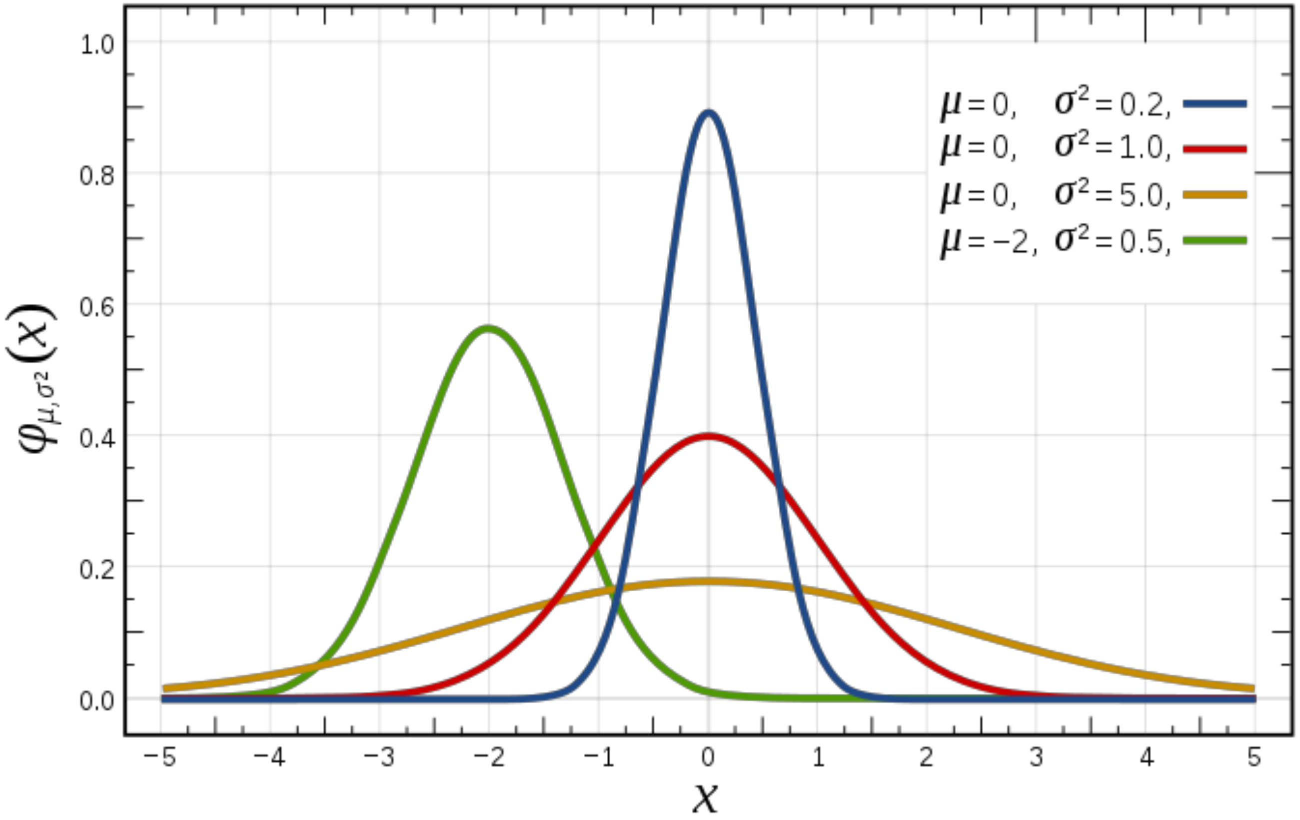
x is a scalar (single number).

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

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

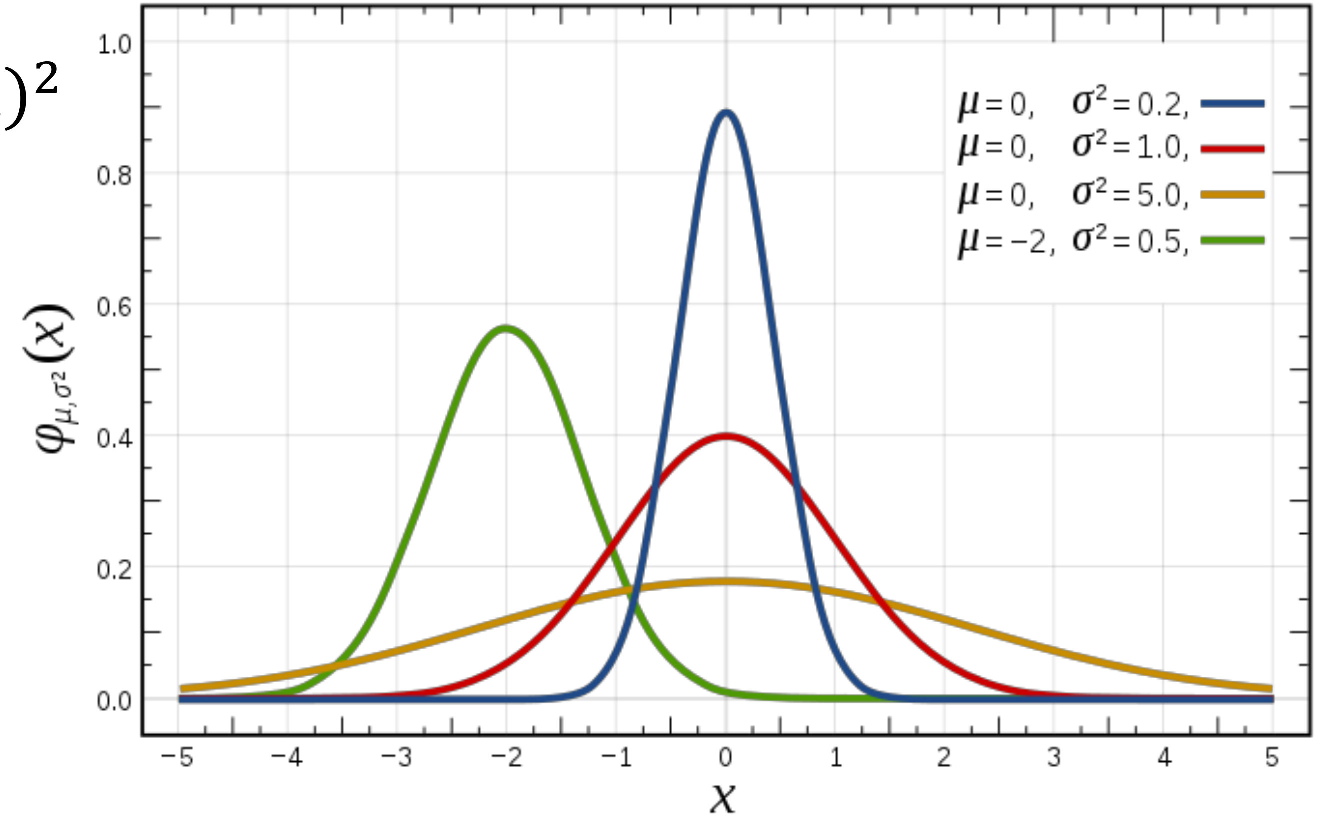
where μ is the mean and σ^2 is the variance.

X ~***N*** (μ, σ)



Variance

Sample Variance: $\sigma^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu)^2$



Tell us how much the data varies around the mean μ.

Low variance produce a narrow curve, and high variance produces a flatter curve.

Multivariate Gaussian

We want to capture the distribution with multiple random variables.

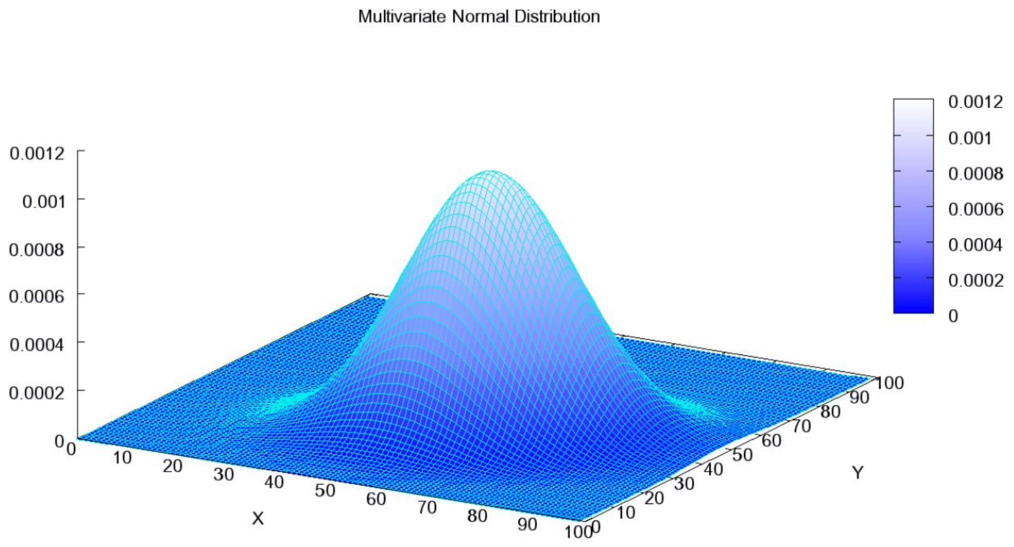
Can't we just use multiple univariate gaussians?

No, because we would lose the variance between the variables.

Thus, *x* (a scalar) now becomes ***X*** (a vector of data points).

For the mean, how does **μ** change now that I have a vector ***X*** ?

μ becomes a vector, each position matching the mean within ***X***.



JAMES MADISON UNIVERSITY

Covariance

Covariance is how the components within *X* are related.

How many covariance values exist?

One between each pair of values. We will encode this in a *n x n* matrix.

$$Cov[X,Y] = \frac{\sum_{i=1}^{N}\left((X_i - \mu_x)\,(Y_i - \mu_y)\right)}{n-1}$$

Interpret $cov(x,y) > 0$ — y tends to increase when x $increases$.

Interpret $cov(x,y) < 0$ — y tends to decrease $when\ x\ decreases$

Interpret $cov(x,y) = 0$ — x and y are independent of each other.

Covariance

$$Cov[X,Y] = \frac{\sum_{i=1}^{N}\left((X_i - \mu_x)\,(Y_i - \mu_y)\right)}{n-1}$$

X = [2.1, 2.5, 3.6, 4.0]

Y = [8, 10, 12, 14]

$Calculate\ the\ mean\ of\ each\ rv.\ \mu_x = \frac{12.2}{4} = 3.05$

$Calculate\ the\ mean\ of\ each\ rv.\ \mu_y = \frac{44}{4} = 11$

COV(X,X) = ((2.1 – 3.05)(2.1 – 3.05) + (2.5 – 3.05)(2.5 – 3.05) + (3.6 – 3.05)(3.6 – 3.05) + (4 – 3.05)(4– 3.05)) / (n-1)

≈ 0.80

COV(Y,Y) = ((8 – 11)(8 – 11) + (10 – 11)(10 – 11) + (12 – 11)(12 – 11) + (14 – 11)(14– 11)) / (n-1)

≈ 0.67

COV(X,Y) = ((2.1 – 3.05)(8 – 11) + (2.5 – 3.05)(10 – 11) + (3.6 – 3.05)(12 – 11) + (4 – 3.05)(14– 11))/(n-1)

≈ 2.27

COV(X,Y) = $\begin{bmatrix} .80 & 2.27 \\ 2.27 & .67 \end{bmatrix}$

Correlation

What is correlation?

You can think of correlation as the normalized covariance (goes from -1, 1).

$$Corr[X,Y] = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

COV(Y,Y) = ((8 – 11)(8 – 11) + (10 – 11)(10 – 11) + (12 – 11)(12 – 11) + (14 – 11)(14– 11)) / (n-1)

≈ 0.67

COV(X,Y) = ((2.1 – 3.05)(8 – 11) + (2.5 – 3.05)(10 – 11) + (3.6 – 3.05)(12 – 11) + (4 – 3.05)(14– 11))/(n-1)

≈ 2.27

COV(X,Y) = $\begin{bmatrix} .80 & 2.27 \\ 2.27 & .67 \end{bmatrix}$

Density for Multivariate

$$f(x) = \frac{1}{\sqrt{|2\pi C|}} \exp\left(-\frac{1}{2}(x - \mu)^T C^{-1}(x - \mu)\right)$$

Where $|2\pi C| = \left((2\pi)^n \det C\right)$

What is C^{-1} ?

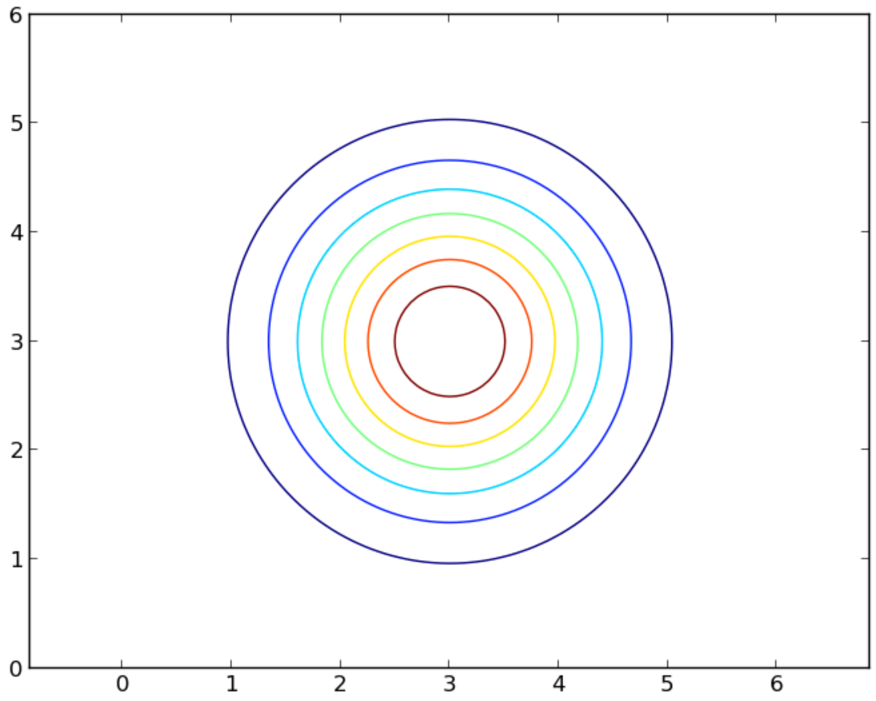
Lets review the formula for the univariate case. $p(x;\, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi q}} \exp\left(-\frac{1}{2\sigma^2}(x - \mu^2)\right)$

We need the variance to be in the denominator. But how do we do that with a matrix?

C^{-1} is the inverse of the matrix C

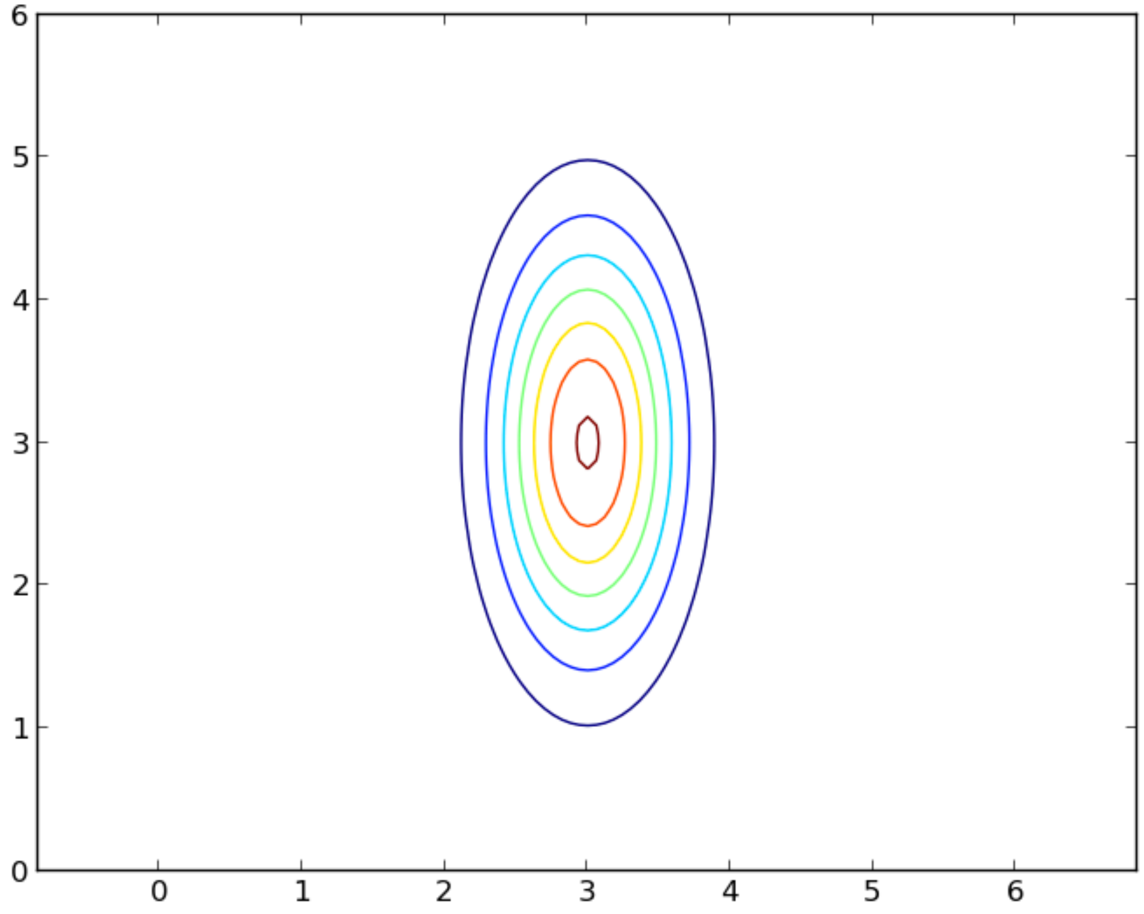
Multivariate PDF Example

$$\mu = \begin{bmatrix} 3 \\ 3 \end{bmatrix} \quad \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$



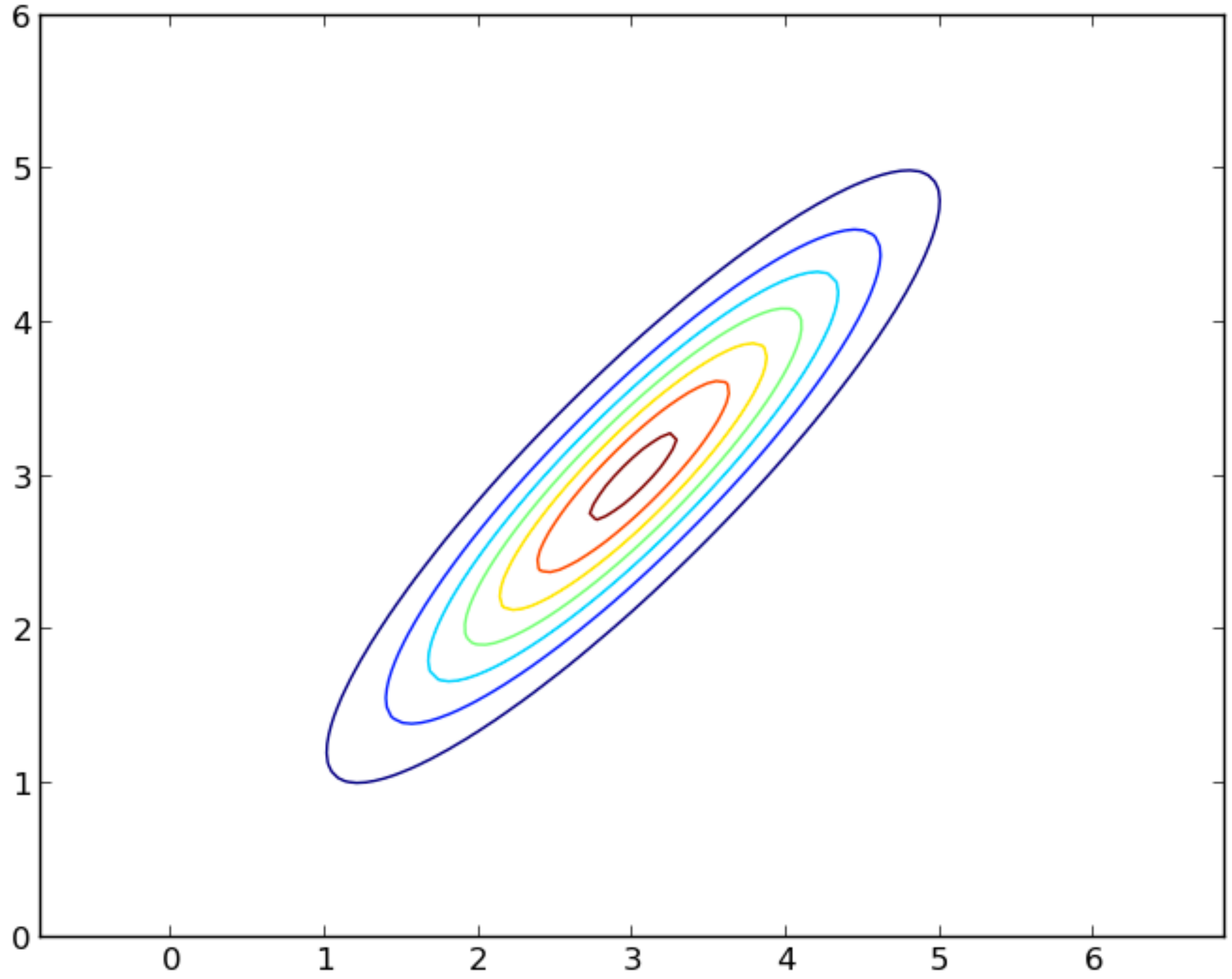
Multivariate PDF Example

$$\mu = \begin{bmatrix} 3 \\ 3 \end{bmatrix} \; \Sigma = \begin{bmatrix} .2 & 0 \\ 0 & 1 \end{bmatrix}$$



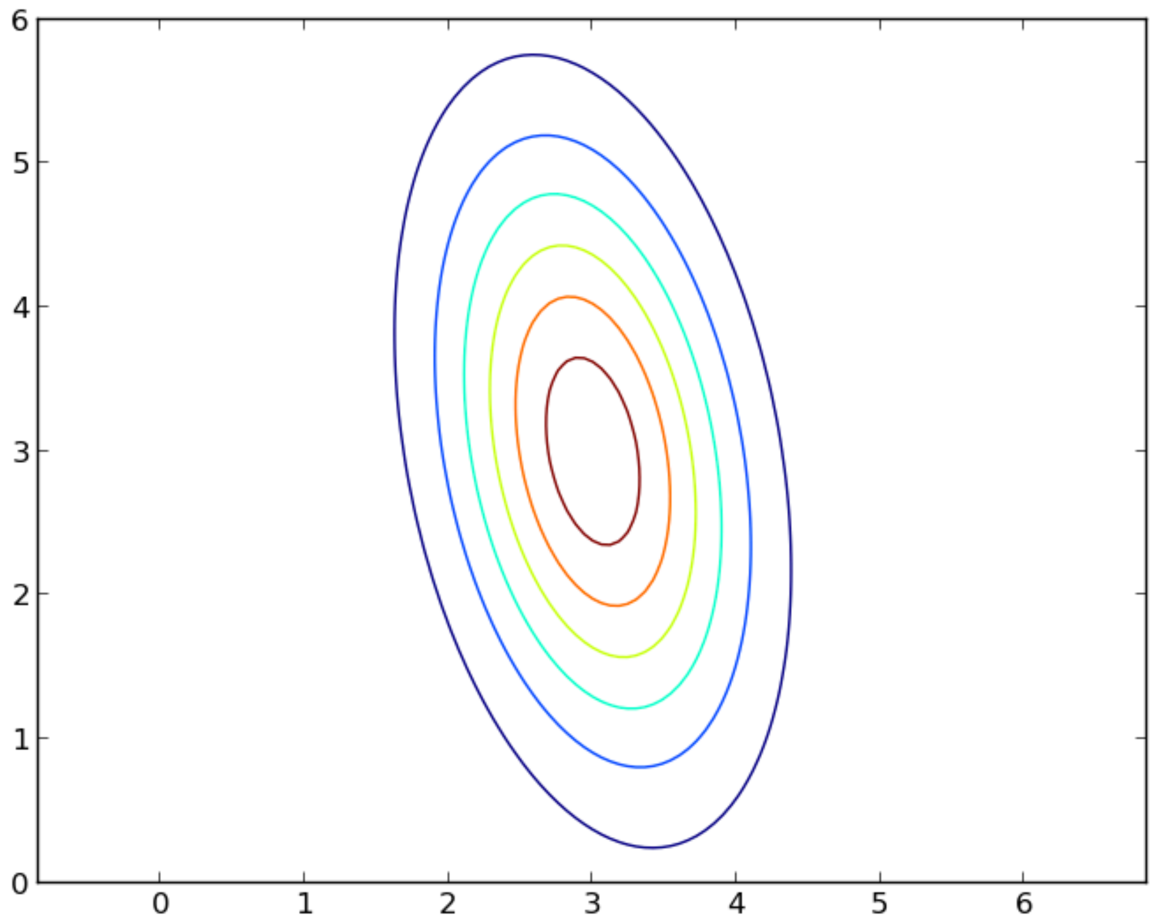
Multivariate PDF Example

$$\mu = \begin{bmatrix} 3 \\ 3 \end{bmatrix} \Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



Multivariate PDF Example

$$\mu = \begin{bmatrix} 3 \\ 3 \end{bmatrix} \; \Sigma = \begin{bmatrix} .5 & -.3 \\ -.3 & 2 \end{bmatrix}$$



Updating Gaussian distributions

Prediction step: if $P(X_t \mid e_{1:t})$ is Gaussian, then prediction

$$P(X_{t+1}|e_{1:t}) = \int_{x_t} P(X_{t+1}|\, x_t)P(x_t|\, e_{1:t})dx_t$$

is Gaussian.

If P(Xt+1 | e1:t) is Gaussian, then the update distribution

$$P(X_{t+1}|\, e_{1:t+1}) = \alpha P(e_{t+1}|\, X_{t+1})P(X_{t+1}|\, e_{1:t})$$

is also Gaussian.