1. Thursday, September 16

1.1. Adding random variables. If \( x \) and \( y \) are independent:
- they are uncorrelated
- \( \text{cov}(x, y) \), also written \( \sigma_{xy}^2 \), is 0
- Notice that in general correlation and covariance are not the same thing

Let \( z = x + y \). Then \( \mu_z = \mu_x + \mu_y \), and in this case \( \sigma_z^2 = \sigma_x^2 + \sigma_y^2 \).

Convolve the functions to get the distribution of \( z \):

\[
p_z(z) = p_x(x) \ast p_y(y) = \int_{-\infty}^{\infty} p_x(x)p_y(z - x) \, dx
\]

1.2. Bayes.

\[
p(x|y) = \frac{p(x, y)}{p(y)} = \frac{p(x, y)}{\int p(x, y) \, dx} = \frac{p(y|x)p(x)}{\int p(y|x)p(x) \, dx}
\]

A common expression:
- \( x \) is a vector of random variables
- \( w_1 \) and \( w_2 \) are classes

\[
P(w_i|x) = \frac{p(x|w_i)P(w_i)}{p(x)}
\]

That is:

\[
(posterior) = \frac{(likelihood)(prior)}{(evidence)}
\]

1.3. Normal distributions.

\[
p(x) \sim N(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}
\]

A standardized random variable: \( y = (x - \mu)/\sigma \). When \( x \) is normal:

\[
p(y) = \frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}} \sim N(0, 1)
\]

\( \text{erf}(y) \) is the area under a zero-mean, unit-variance Gaussian between \(-\sqrt{2}y\) and \(\sqrt{2}y\)

\[
\text{erf}(y) = \frac{2}{\sqrt{\pi}} \int_{0}^{y} e^{-t^2} \, dt
\]

1.4. Making things Gaussian. You need to find:

\[
\int \alpha e^{-(x^2+bx+c)} \, dx
\]

Complete the square:

\[
\int \alpha e^{-(x^2+bx+c)} \, dx = \int \alpha e^{-(x^2+bx+b^2/4-b^2/4+c)} \, dx
\]

Let \( b = -2\mu \), so \( \mu^2 = b^2/4 \).
\[ \int \alpha e^{-\frac{(x+b/2)^2}{2}} e^{-\frac{b^2}{4+c}} \, dx = \alpha e^{-\frac{b^2}{4+c}} \int e^{-\frac{(x+b/2)^2}{2}} \, dx \]

The integral becomes \( \int \sqrt{\pi} N(-b/2, 1/2) \, dx = \sqrt{\pi} \), so you get:

\[ \int \alpha e^{-\frac{(x^2+b^2x+c}{2}} \, dx = \alpha \sqrt{\pi} e^{-\frac{(b^2}{4+c}} \]

1.5. **Multivariate Gaussian.** When \( x = [x_1, x_2, ..., x_d] \):

- \( \mu = E[x] \) (the mean vector)
- \( \sigma_i^2 = E[(x_i - \mu_i)^2] \)
- \( \sigma_{ij}^2 = E[(x_i - \mu_i)(x_j - \mu_j)] \)

This last one gives you the covariance matrix:

\[ \Sigma = E[(x - \mu)(x - \mu)^T] \]

1.6. **Linear transformations of Gaussian random variables.** Let \( y = A^T x \). \( A \) is a \( k \times d \) matrix, \( x \) is a \( d \)-length vector of rand vars with \( p(x) \sim N(\mu, \Sigma) \).

Mean of \( y \) is \( A^T \mu \). Covariance is \( A^T xx^TA = A^T \Sigma A \).

So \( p(y) \sim N(A^T \mu, A^T \Sigma A) \).

1.7. **Sums of Gaussian distributions.** Sums of either dependent or independent Gaussian rand vars are Gaussian rand vars.

Suppose \( x_1 \ldots x_d \) are independent variables, with each \( x_i \) having \( p_x(x) \sim N(\mu_i, \sigma_i^2) \).

\[ p(x) = \prod_i p(x_i) = \frac{1}{(2\pi)^{d/2}} \prod_i \sigma_i \exp \left[ -\frac{1}{2} \sum_i \left( \frac{x_i - \mu_i}{\sigma_i} \right)^2 \right] \]

\[ p(x) = \frac{1}{(2\pi)^{d/2} \sqrt{|\Sigma|}} \exp \left[ -\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right] \]

1.8. **Mahalanobis distance.** Define \( r = (x - \mu)^T \Sigma^{-1} (x - \mu) \), in the thing above. What’s nice is that \( p(r) \sim N(0, I) \).

Setting \( r = 1 \) gives you an ellipse in a standardized place around a gaussian. Its center/maximum is at the mean. This is the ellipse at Mahalanobis distance 1 from the mean.