WTF is a multivariate normal?

Pete Bunch

Lent Term 2012

1 Introduction

Normal (or Gaussian) distributions are the bread and butter of 6th form stats and first year probability. Now all of a sudden we've got them in multiple dimensions. What is it and where did it come from?

2 Univariate Normal

A univariate normal probability density function (pdf) looks like this,

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

Lets break this down a bit. First, the bit in front of the exponential is just a normalising factor. It's there simply to ensure that when we integrate the pdf we get 1. Lets just call this Z for now, and we'll worry about what it is later.

Now, the distinctive shape of the density function is determined by the fact that its an exponential of something squared. The "something squared" is just the squared scaled distance,

$$d^2 = \left(\frac{x-\mu}{\sigma}\right)^2. \tag{2}$$

This is the square of the distance from the mean, scaled by σ, which is just a length-scale parameter. So our pdf really just boils down to,

$$p(x) = Z \exp\left\{-\frac{1}{2}d^2\right\}. \tag{3}$$

3 Adding dimensions

We can extend the univariate Gaussian to more dimensions by just redefining d^2 as a function of multiple coordinates, multiple means and multiple length scales. For example, in 3D we might want the density at a point (x, y, z), in which case we could use,

$$d^2 = \left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 + \left(\frac{z-\mu_z}{\sigma_z}\right)^2. \tag{4}$$

This is then plugged into (3) to find the density value at this point.

In more than one dimension, multiple points will have the same probability density. If we fix the distance at a particular value, $d = k$, then we get the equation of an ellipse (in 2D), an ellipsoid (in 3D) or a hyper-ellipsoid (in > 3D).

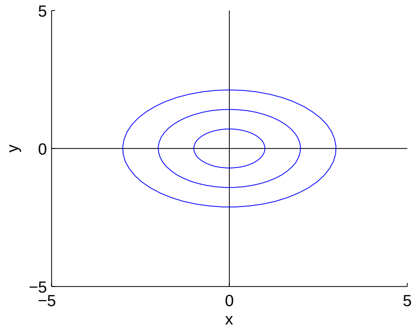

Figure 1: Constant density ellipses at $d = 1$, $d = 2$, $d = 3$, for a zero-mean normal distribution with $\sigma_x = 0.5$, $\sigma_y = 1$.

We can write (4) more concisely in vector/matrix notation as,

$$\begin{aligned}
d^2 &= \begin{bmatrix} x-\mu_x & y-\mu_y & z-\mu_z \end{bmatrix} \begin{bmatrix} \sigma_x^2 & 0 & 0 \\ 0 & \sigma_y^2 & 0 \\ 0 & 0 & \sigma_z^2 \end{bmatrix}^{-1} \begin{bmatrix} x-\mu_x \\ y-\mu_y \\ z-\mu_z \end{bmatrix} \\
&= \left(\begin{bmatrix} x \\ y \\ z \end{bmatrix} - \begin{bmatrix} \mu_x \\ \mu_y \\ \mu_z \end{bmatrix} \right)^T \begin{bmatrix} \sigma_x^2 & 0 & 0 \\ 0 & \sigma_y^2 & 0 \\ 0 & 0 & \sigma_z^2 \end{bmatrix}^{-1} \left(\begin{bmatrix} x \\ y \\ z \end{bmatrix} - \begin{bmatrix} \mu_x \\ \mu_y \\ \mu_z \end{bmatrix} \right) \\
&= (\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Lambda}^{-1} (\mathbf{x}-\boldsymbol{\mu}). \qquad (5)
\end{aligned}$$

Good. We're nearly done. The only problem is that the constant-density ellipses we get using this method are always aligned with the axes. What if we wanted there to be a correlation between variables? Instead of starting again with a new (and more complicated) distance metric, we can solve this problem geometrically. We can get any ellipse by first choosing σ_x, σ_y, etc. to give us the right scale, and then rotating them to give the right orientation. In fact, its the $(\mathbf{x}-\boldsymbol{\mu})$ factors that need to be rotated, so our new, generalised distance metric will be,

$$\begin{aligned}
d^2 &= (\mathbf{Q}(\mathbf{x}-\boldsymbol{\mu}))^T \boldsymbol{\Lambda}^{-1} (\mathbf{Q}(\mathbf{x}-\boldsymbol{\mu})) \\
&= (\mathbf{x}-\boldsymbol{\mu})^T \left(\mathbf{Q}^T \boldsymbol{\Lambda}^{-1} \mathbf{Q}\right) (\mathbf{x}-\boldsymbol{\mu}) \\
&= (\mathbf{x}-\boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}-\boldsymbol{\mu}) \\
&. \qquad (6)
\end{aligned}$$

This is called the squared Mahalanobis distance.

$\mathbf{Q}$ is a rotation matrix, which means it's orthonormal, i.e. $\mathbf{Q}^{-1} = \mathbf{Q}^T$. This means that the new covariance matrix is,

$$\Sigma = \mathbf{Q}^T \boldsymbol{\Lambda} \mathbf{Q} \qquad (7)$$

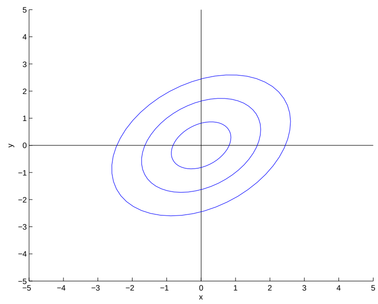

Figure 2: Constant density ellipses at $d = 1$, $d = 2$, $d = 3$, for the same normal distribution rotated by 45°

This looks like an eigen-decomposition, because it is. Remember that a symmetric matrix always has orthogonal eigenvectors? The converse is also true, so the covariance matrix of a multivariate Gaussian is always symmetric. Also, because the σs have to be positive, the matrix must be positive-definite.

Finally, lets put it all together. Substituting the Mahalanobis distance into (3), we get,

$$p(x) = Z \exp\left\{-\frac{1}{2}\left[(\mathbf{x}-\boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}-\boldsymbol{\mu})\right]\right\}. \tag{8}$$

Integrating the whole thing, we find that $Z = |2\pi\mathbf{\Sigma}|^{-1/2}$ (This can also be reached through logical and geometric extensions of the univariate case).

$$p(x) = |2\pi\mathbf{\Sigma}|^{-1/2} \exp\left\{-\frac{1}{2}\left[(\mathbf{x}-\boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}-\boldsymbol{\mu})\right]\right\}. \tag{9}$$