Lecture 9, Feb 5, 2024

Joint Gaussian Distributions

Theorem

Central Limit Theorem: Let $X_1, \ldots, X_n$ be a sequence of i.i.d. RVs from any distribution with finite mean μ and variance σ^2, and let $S_n = X_1 + \cdots + X_n$ be their sum; and let

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

which is zero-mean and unit variance, then

$$\lim_{n\to\infty} P[Z_n \leq z] = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} e^{-\frac{x^2}{2}}\,\mathrm{d}x$$

i.e. the distribution of Z_n approaches $\mathcal{N}(0,1)$.

Definition

X and Y are *jointly Gaussian* if their joint PDF is given by

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho_{X,Y}^2}} e^{-\frac{1}{2(1-\rho_{X,Y}^2)}\left(\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho_{X,Y}\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right)}$$

where μ_1, μ_2 are the means, σ_1^2, σ_2^2 are the variances, and $\rho_{X,Y}$ is the correlation coefficient of X, Y.

- Notice that the expression is symmetric in X and Y, and both variables appear in their normalized form
- If X and Y are uncorrelated, then $f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{1}{2}\left(\left(\frac{x-\mu_1}{\sigma_1}\right)^2 + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right)} = f_X(x)f_Y(y)$
  - For Gaussian RVs, uncorrelated implies independent
- If we compute marginals by completing the square, we see that both are Gaussian

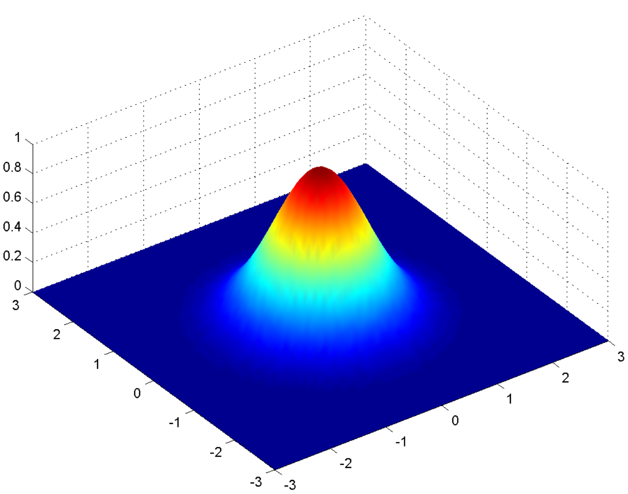

Figure 1: Plot of a joint Gaussian distribution with zero-mean, unit variance and uncorrelated X, Y.

- For the case of zero-mean, unit variance and uncorrelated X, Y above the contours of constant probability are circles centered about the origin

- Changing the mean shifts the centre of the distribution
- The exponent is in quadratic form
- If the variances are not equal (but still uncorrelated), we will get axis-aligned ellipses as the distribution in each dimension gets stretched out
- If the correlation is nonzero, the axes of the ellipse will no longer be axis-aligned
  * For a positive ρ the ellipse is along the $x = y$ axis
  * For a negative ρ the ellipse is along the $x = -y$ axis
  * The closer ρ is to 1, the more tightly packed the ellipse is along its axis
- We can always find a transformation that aligns the axes of the ellipse with the x and y axis to make them independent in the new transformed space

- The conditional PDF is $f_{X,Y}(x|y) = \frac{1}{\sqrt{2\pi\sigma_1^2(1-\rho_{X,Y}^2)}}e^{-\frac{1}{2(1-\rho_{X,Y}^2)\sigma_1^2}\left(x-\rho_{X,Y}\frac{\sigma_1}{\sigma_2}(y-\mu_2)-\mu_1\right)^2}$
  - This is another Gaussian with $\mu = \rho_{X,Y}\dfrac{\sigma_1}{\sigma_2}(y-\mu_2)+\mu_1$ and $\sigma^2 = (1-\rho_{X,Y}^2)\sigma_1^2$
  - Notice the new mean is the normalized y, scaled up by the standard deviation of x, multiplied by the correlation and then add back to mean of x to shift it
  - The variance has no dependence on y but knowing y reduces the variance of x
  - As $\rho_{X,Y} \to \pm 1$, the conditional variance approaches 0 because X is just a linear function of Y
- Consider a linear transformation $\begin{bmatrix} V \\ W \end{bmatrix} = \begin{bmatrix} a & b \\ c & e \end{bmatrix}\begin{bmatrix} X \\ Y \end{bmatrix} = \boldsymbol{A}\boldsymbol{X}$ where the determinant is nonzero (invertible)
  - The joint PDF of V and W is given by $f_{V,W}(v,w) = \dfrac{f_{X,Y}(x,y)}{\det \boldsymbol{A}}$
  - Intuitively an area $\mathrm{d}x$ by $\mathrm{d}y$ is mapped to an area of size $\mathrm{d}P$; this ratio is the determinant
    * $f(x,y)\,\mathrm{d}x\,\mathrm{d}y = f(v,w)\,\mathrm{d}P$ since both are the probability of a small region
  - For a nonlinear transformation the determinant is replaced by a Jacobian
  - Note practically to get this in terms of v, w we need to find the inverse mapping from v, w to x, y
- More generally consider $\boldsymbol{Z} = \boldsymbol{A}\boldsymbol{Z}$ where $A \in \mathbb{R}^{n\times n}$ and is invertible
  - The joint PDF is $f_Z(\boldsymbol{Z}) = f(z_1, \dots, z_n) = \dfrac{f(x_1, \dots, x_n)}{\det \boldsymbol{A}} = \dfrac{f_{\boldsymbol{X}}(\boldsymbol{A}^{-1}\boldsymbol{z})}{\det \boldsymbol{A}}$

Generalization of Expectation and Variance

- Let the *mean vector* of $\boldsymbol{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}$ be $\boldsymbol{m_X} = E[\boldsymbol{X}] = \begin{bmatrix} E[X_1] \\ \vdots \\ E[X_n] \end{bmatrix}$
- Let the *correlation matrix* be $\boldsymbol{R_X} = \begin{bmatrix} E[X_1^2] & E[X_1X_2] & \dots & E[X_1X_n] \\ E[X_2X_1] & E[X_2^2] & \dots & E[X_2X_n] \\ \vdots & \vdots & \ddots & \vdots \\ E[X_nX_1] & E[X_nX_2] & \dots & E[X_n^2] \end{bmatrix}$
  - Note that this is symmetric
  - The diagonal elements are second moments
- Let the *covariance matrix* be $\boldsymbol{K_X}$ such that entry (i,j) is σ_{X_i,X_j}
  - This is symmetric positive semidefinite
  - The diagonal entries are the variances of each variable
  - If the means are all zero, this is equivalent to the correlation matrix
  - If all X_i, X_j are uncorrelated, then the covariance matrix is diagonal
- Notice that $\boldsymbol{R_X} = E[\boldsymbol{X}\boldsymbol{X}^T]$ and $\boldsymbol{K_X} = E[(\boldsymbol{X}-\boldsymbol{m_X})(\boldsymbol{X}-\boldsymbol{m_X})^T] = \boldsymbol{R_X} - \boldsymbol{m_X}\boldsymbol{m_X^T}$
- For any general linear transformation $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X}$:
  - $E[\boldsymbol{Y}] = \boldsymbol{A}E[\boldsymbol{X}] = \boldsymbol{A}\boldsymbol{m_X}$
  - $\boldsymbol{K_Y} = \boldsymbol{A}\boldsymbol{K_X}\boldsymbol{A}^T$
- We can apply an eigendecomposition to the covariance matrix
  - Often our covariance matrix will be full rank, which makes it positive definite, and makes the

decomposition always possible

- Find eigenvectors $\boldsymbol{e}_i$ such that $\boldsymbol{K}_{\boldsymbol{X}}\boldsymbol{e}_i = \lambda \boldsymbol{e}_i$ and $\boldsymbol{e}_i^T \boldsymbol{e}_j = \delta_{ij}$ (orthonormal eigenvectors)
- Let $\boldsymbol{P} = \begin{bmatrix} \boldsymbol{e}_1 & \dots & \boldsymbol{e}_n \end{bmatrix}$ and $\boldsymbol{\Lambda} = \operatorname{diag} \lambda_i$, then $\boldsymbol{K}_{\boldsymbol{X}} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^T$
- For a general Gaussian, this means that if we first transform the variables by $\boldsymbol{P}^T$, then they will all be independent of each other