The multivariate Gaussian

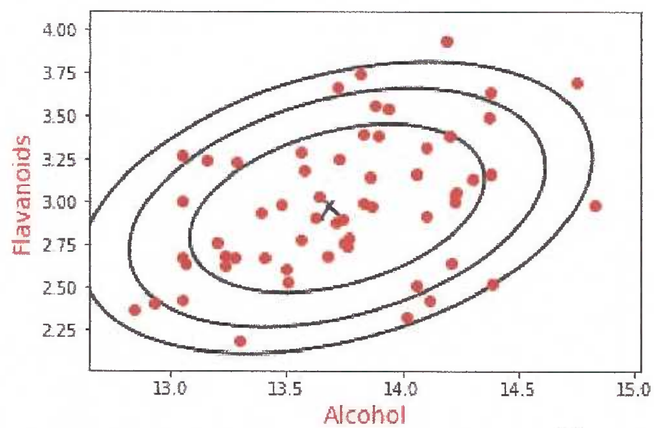
POPULAR & POWERFUL PROBABILITY
DISTRIBUTION FOR DATA OF
ARBITRARY DIMENSION

Topics we'll cover

1. Functional form of the density
2. Special case: diagonal Gaussian
3. Special case: spherical Gaussian
4. Fitting a Gaussian to data

Recall: the bivariate Gaussian

$X_2 \uparrow$



$\rightarrow X_1$

Bivariate Gaussian, parametrized by:

$$\text{mean } \mu = \begin{pmatrix} \underbrace{13.7}_{\mu \text{ of } X_1} \\ 3.0 \end{pmatrix} \text{ and covariance matrix } \Sigma = \begin{pmatrix} \overbrace{0.20}^{Var(X_1)} & \overbrace{0.06}^{Cov(X_1, X_2)} \\ 0.06 & \underbrace{0.12}_{Var(X_2)} \end{pmatrix}$$

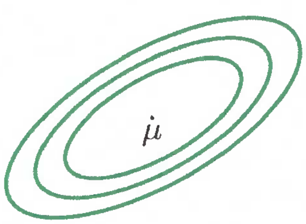
μ of X_2 or μ_{X_2}

Cov(X_1, X_2) positive correlation between the 2 features coz it's tilted upward

computed the variance of each feature & covariance between features

$$\Sigma = \begin{bmatrix} Var(X_1) & Cov(X_1,X_2) & Cov(X_1,X_3) \\ & \ddots & \\ & & Var(X_d) \end{bmatrix}$$

The multivariate Gaussian



$N(\mu, \Sigma)$: Gaussian in $\mathbb{R}^d$

- mean: $\mu \in \mathbb{R}^d$
- covariance: $d \times d$ matrix Σ

Generates points $X = (X_1, X_2, \ldots, X_d)$.

- μ is the vector of coordinate-wise means:

$$\mu_1 = \mathbb{E}X_1, \ \mu_2 = \mathbb{E}X_2, \ldots, \ \mu_d = \mathbb{E}X_d.$$

- Σ is a matrix containing all pairwise covariances:

$$\Sigma_{ij} = \Sigma_{ji} = \text{cov}(X_i, X_j) \quad \text{if } i \neq j$$
$$\Sigma_{ii} = \text{var}(X_i)$$

Formula for

→ Density $$p(x) = \boxed{\frac{1}{(2\pi)^{d/2}\underbrace{|\Sigma|^{1/2}}_{\text{determinant}}}}_{\text{normalizes}} \exp\left(-\frac{1}{2}\underbrace{(x-\mu)^T\Sigma^{-1}(x-\mu)}_{\text{most important that depends on } X}\right)$$

reduces to $\boxed{x^T M x}$ if $\mu = \emptyset$

ellipsoid that is kind of quadratic

Special case: independent features

Indep'd t.f. uncorrelated so Cov = 0

$\mathrm{Cov}(x_i, x_j) = 0$

Suppose the X_i are independent, and $\mathrm{var}(X_i) = \sigma_i^2$.

What is the covariance matrix Σ, and what is its inverse Σ^{-1}?

$$\Sigma = \begin{bmatrix} \mathrm{Var}(x_1) & & 0 \\ & \ddots & \\ 0 & & \mathrm{Var}(x_d) \end{bmatrix}$$

Diagonal Matrix

$$= \begin{bmatrix} \sigma_1^2 & & & 0 \\ & \sigma_2^2 & & \\ & & \ddots & \\ 0 & & & \sigma_d^2 \end{bmatrix} = \mathrm{diag}(\sigma_1^2 \cdots \sigma_d^2)$$

Inverse → get the inverse of diagonal

$$\Sigma^{-1} = \begin{bmatrix} 1/\sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & 1/\sigma_d \end{bmatrix}$$

Diagonal Gaussian

Number of Params

$d + d = 2d$ params

Diagonal Gaussian: the X_i are independent, with variances σ_i^2. Thus

$$\Sigma = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_d^2) \quad \text{(off-diagonal elements zero)}$$

No correlation

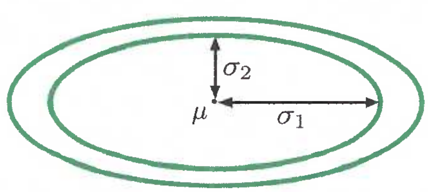
axis aligned

Each X_i is an independent one-dimensional Gaussian $N(\mu_i, \sigma_i^2)$:

Density simplifies

$$\Pr(x) = \Pr(x_1)\Pr(x_2)\cdots\Pr(x_d) = \underbrace{\frac{1}{(2\pi)^{d/2}\sigma_1 \cdots \sigma_d} \exp\left(-\sum_{i=1}^{d} \frac{(x_i - \mu_i)^2}{2\sigma_i^2}\right)}$$

Contours of equal density are **axis-aligned ellipsoids** centered at μ:



Even more special case: spherical Gaussian

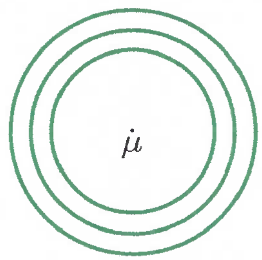
The X_i are independent and all have the same variance σ^2.

$$\longrightarrow \Sigma = \sigma^2 I_d = \underbrace{\text{diag}(\sigma^2, \sigma^2, \ldots, \sigma^2)} \quad \text{(diagonal elements } \sigma^2 \text{, rest zero)}$$

Each X_i is an independent univariate Gaussian $N(\mu_i, \sigma^2)$:

$$\Pr(x) = \Pr(x_1)\Pr(x_2)\cdots\Pr(x_d) = \underbrace{\frac{1}{(2\pi)^{d/2}\sigma^d} \exp\left(-\frac{\|x-\mu\|^2}{2\sigma^2}\right)}$$

Density at a point depends only on its distance from μ:



popular gaussian

☆☆☆ Concrete & Practical

How to fit a Gaussian to data

→ Compute the mean and covariance of these data

Fit a Gaussian to data points $x^{(1)}, \ldots, x^{(m)} \in \mathbb{R}^d$. μ and Σ(sigma) are parameters of Gaussian

- Empirical mean $\rightsquigarrow \mu$

$$\mu = \frac{1}{m}\left(x^{(1)} + \cdots + x^{(m)}\right)$$

→ add the m data points & divide by m

Notation: $Cov(X_i, X_j) = Cov(X, Y)$

- Empirical covariance matrix has i, j entry: apply formula for covariance

$$\boxed{\Sigma_{ij} = \left(\frac{1}{m}\sum_{k=1}^{m} x_i^{(k)} x_j^{(k)}\right) - \mu_i\mu_j}$$

$E[X_i X_j] = \frac{1}{m}\sum_{k=1}^{m} x_i^{(k)} x_j^{(k)}$

$Cov(X_i, X_j) \equiv E[X_i X_j] - E[X_i]E[X_j]$

Since we have computed μ_i and μ_j, so we compute the average value of X_i times X_j in the dataset.