Two-dimensional generative modeling with the bivariate Gaussian

Topics we'll cover

1. Generative modeling of two-dimensional data
2. The bivariate Gaussian distribution
3. Decision boundary of the generative model

The winery prediction problem

Which winery is it from, 1, 2, or 3?

Using one feature ('Alcohol'), error rate is 29%.

What if we use **two** features?

The data set, again

Training set obtained from 130 bottles

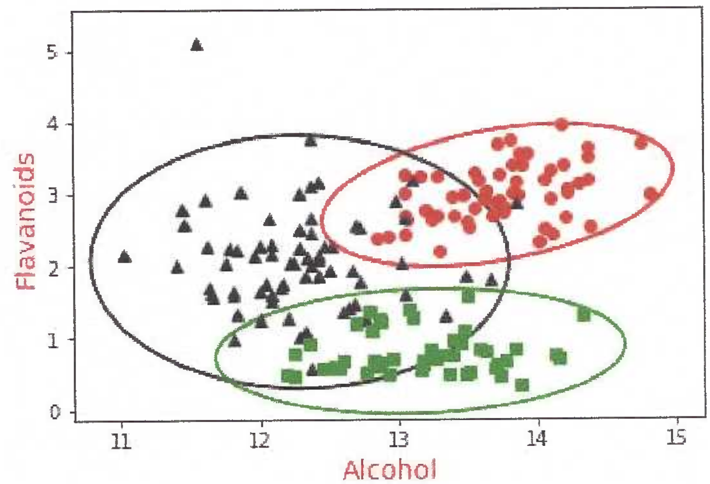
- Winery 1: 43 bottles
- Winery 2: 51 bottles
- Winery 3: 36 bottles
- For each bottle, 13 features:
  'Alcohol', 'Malic acid', 'Ash', 'Alcalinity of ash','Magnesium',
  'Total phenols', 'Flavanoids', 'Nonflavanoid phenols', 'Proanthocyanins',
  'Color intensity', 'Hue', 'OD280/OD315 of diluted wines', 'Proline'

Also, a separate test set of 48 labeled points.

This time: 'Alcohol' and 'Flavanoids'.

Why it helps to add features

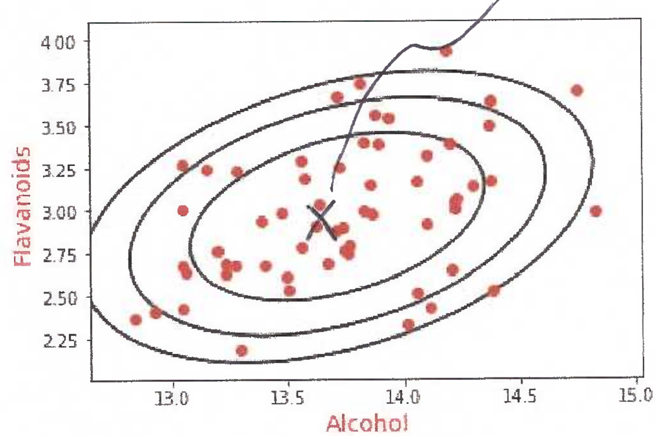
Better **separation** between the classes!



Error rate drops from 29% to 8%.

The bivariate Gaussian

mean



Covariance between alcohol & flavan

Model class 1 by a bivariate Gaussian, parametrized by:

$$\text{mean } \mu = \begin{pmatrix} 13.7 \\ 3.0 \end{pmatrix} \text{ and covariance matrix } \Sigma = \begin{pmatrix} 0.20 & 0.06 \\ 0.06 & 0.12 \end{pmatrix}$$

variance

Dependence between two random variables

Suppose X_1 has mean μ_1 and X_2 has mean μ_2.

Can measure dependence between them by their **covariance**:

- $\text{cov}(X_1, X_2) = \mathbb{E}[(X_1 - \mu_1)(X_2 - \mu_2)] = \mathbb{E}[X_1 X_2] - \mu_1 \mu_2$
- Maximized when $X_1 = X_2$, in which case it is $\text{var}(X_1)$.
- It is at most $\text{std}(X_1)\text{std}(X_2)$.

The bivariate (2-d) Gaussian

A distribution over $(x_1, x_2) \in \mathbb{R}^2$, parametrized by:

- **Mean** $(\mu_1, \mu_2) \in \mathbb{R}^2$, where $\mu_1 = \mathbb{E}(X_1)$ and $\mu_2 = \mathbb{E}(X_2)$
- **Covariance matrix** $\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$ where $\left\{ \begin{array}{c} \Sigma_{11} = \text{var}(X_1) \\ \Sigma_{22} = \text{var}(X_2) \\ \Sigma_{12} = \Sigma_{21} = \text{cov}(X_1, X_2) \end{array} \right\}$

Density is highest at the mean, falls off in ellipsoidal contours.

→ shape of these ellipsoids
ellipsoid contours
is given by
sigma matrix

Density of the bivariate Gaussian

parameters

- **Mean** $(\mu_1, \mu_2) \in \mathbb{R}^2$, where $\mu_1 = \mathbb{E}(X_1)$ and $\mu_2 = \mathbb{E}(X_2)$
- **Covariance matrix** $\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$

$$\text{Density} \quad p(x_1, x_2) = \frac{1}{2\pi \underbrace{|\Sigma|^{1/2}}_{|\Sigma| = \text{determinant of } \Sigma}} \exp\left(-\frac{1}{2} \underbrace{\begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix}^T}_{\substack{X-\mu \\ 1\times 2 \text{ transpose}}} \underbrace{\Sigma^{-1}}_{\substack{2\times 2 \\ \text{matrix}}} \underbrace{\begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix}}_{X-\mu} \right)$$

$X - \mu$ 2×1 vector

$\mu = (\mu_1, \mu_2)$

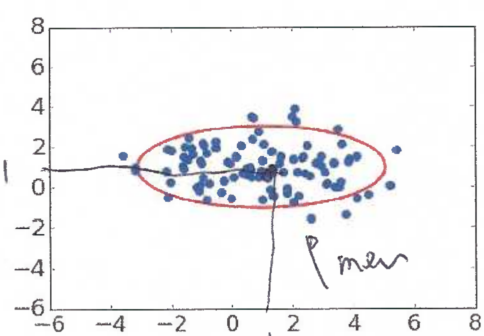
displacement of X from μ

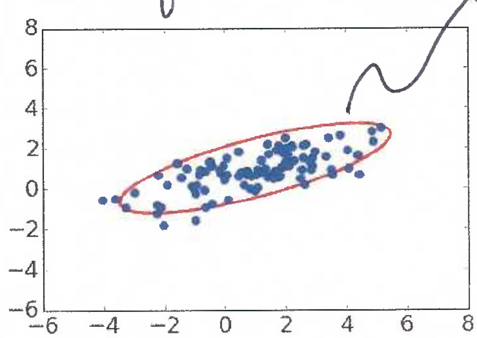
Bivariate Gaussian: examples

In either case, the mean is $(1, 1)$.

Same mean but 2 different cov matrix

representation contour line





Variance X_1

so std $= \sqrt{\text{Var}} = \sqrt{4}$ $\quad \Sigma = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}$ Variance X_2

std$(X_1) = 2$

std$(X_2) = 1$

cov$(X_1, X_2) = 0$

uncorrelated

axis-aligned

$\Sigma = \begin{bmatrix} 4 & 1.5 \\ 1.5 & 1 \end{bmatrix}$ covariance matrix

std$(X_1) = 2$

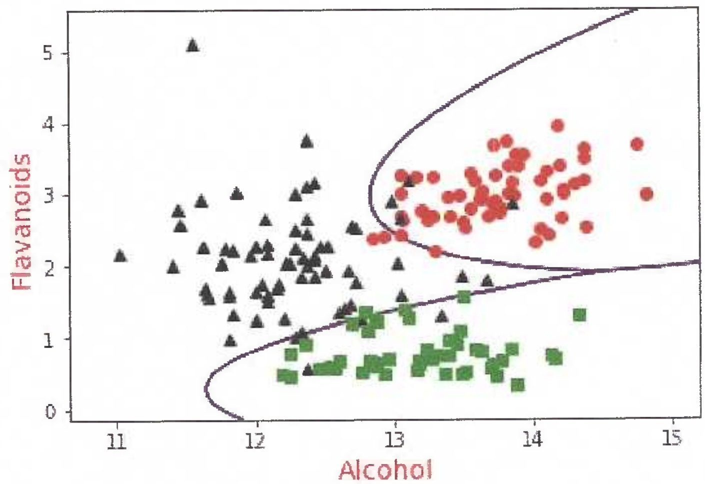
std$(X_2) = 1$

Cov$(X_1, X_2) = 1.5$

correlated

The decision boundary

Go from 1 to 2 features: error rate goes from 29% to 8%.



What kind of function is this? And, can we use more features?