Lesson 7

Bayes Classifier

Awareness through analysis and synthesis

Independent Random Variables

Two random variables X and Y are independent if

$P(x, y) = P(x)P(y)$ for all x in X and y in Y.

Example

Pick a card from the standard deck of playing cards.

X : the suit (That is, spade, heart, and so on)

Y : value (That is, A, K, Q, J, 10, 9, …, 2)

$P(x) = 1/4$, $P(y) = 1/13$ and $P(x, y) = 1/52 = P(x)P(y)$

Independent Random Variables

Throw a fair coin 3 times.

X : number of heads

Y : 1 if last throw is head and 0 otherwise.

There are 8 cases when a coin is thrown 3 times.

HHH, HHT, HTH, HTT, THH, THT, TTH, TTT,

$P(X = 0) = 1/8$, $P(X = 1) = 3/8$, $P(X = 2) = 3/8$ and $P(X= 3) = 1/8$.

$P(Y = 1) = 1/2$ and $P(Y = 0) = 1/2$.

$P(X = 0, Y = 1) = 0 \neq 1/16 = P(X = 0)P(Y = 1)$

X and Y are NOT independent random variables.

Independent Random Variables

Range(X) = {1, 2, 3}

Range(Y) = {1, 2, 3}

			Y	
		1	**2**	**3**
	1	**0.4**	**0.16**	**0.24**
X	**2**	**0.05**	**0.02**	**0.03**
	3	**0.05**	**0.02**	**0.03**

$P(X = 1) = 0.8$, $P(X = 2) = 0.1$, $P(X = 3) = 0.1$

$P(Y = 1) = 0.5$, $P(Y = 2) = 0.2$, $P(Y = 3) = 0.3$

$P(x, y) = P(x)P(y)$

Dependence

Example: Pick a person at random, and take

$$H = \text{height}$$
$$W = \text{weight}$$

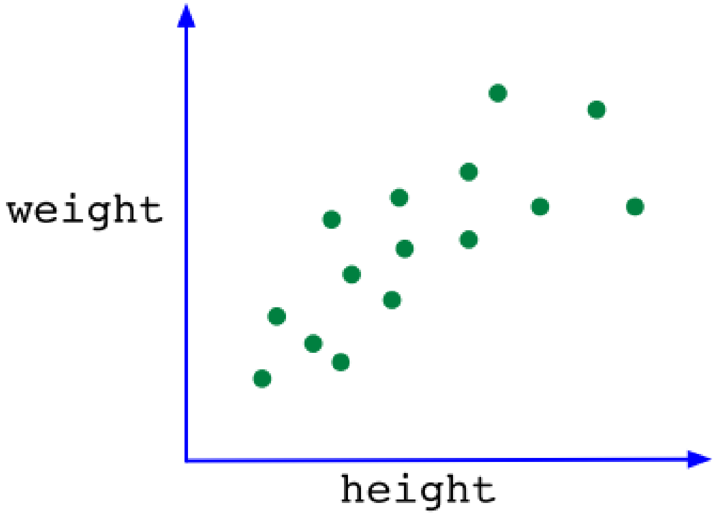
Independence would mean

$$\Pr(H = h, W = w) \;=\; \Pr(H = h)\Pr(W = w).$$

Not accurate: height and weight will be **positively correlated**.

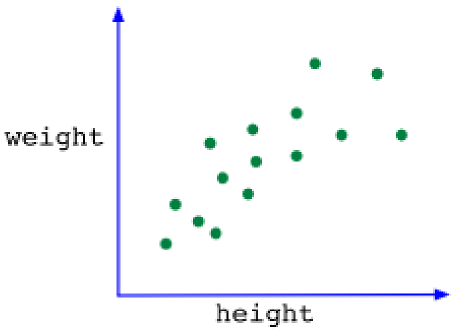
Positive Correlation

H, W are **positively correlated**



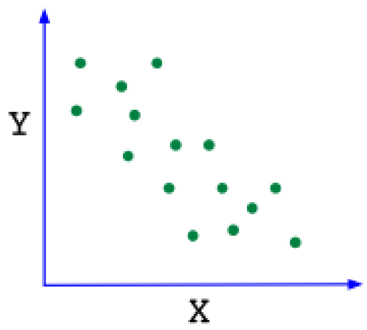
This also implies $\mathbb{E}[HW] > \mathbb{E}[H]\,\mathbb{E}[W]$.

Types of Correlation

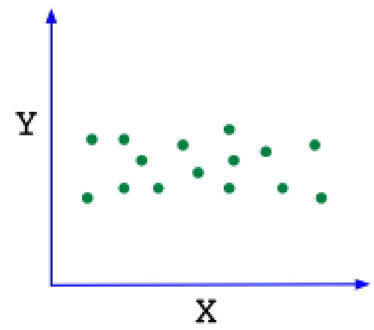


H, *W* **positively correlated**
This also implies

$$\mathbb{E}[HW] > \mathbb{E}[H]\,\mathbb{E}[W]$$



X, *Y* **negatively correlated**
$\mathbb{E}[XY] < \mathbb{E}[X]\,\mathbb{E}[Y]$



X, *Y* **uncorrelated**
$\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$

Covariance and Correlation

- Covariance

$$\begin{aligned}\mathrm{cov}(X, Y) &= \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]\end{aligned}$$

  Maximized when $X = Y$, in which case it is $\mathrm{var}(X)$.
  In general, it is at most $\mathrm{std}(X)\mathrm{std}(Y)$.

- Correlation

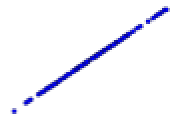
$$\mathrm{corr}(X, Y) = \frac{\mathrm{cov}(X, Y)}{\mathrm{std}(X)\mathrm{std}(Y)}$$

  This is always in the range $[-1, 1]$.

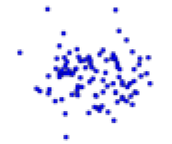
If X, Y independent then $\mathrm{cov}(X, Y) = 0$.
But the converse need not be true.

Correlation Coefficient

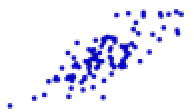
$r = 1$



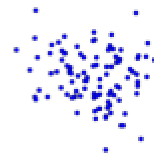
$r = 0$



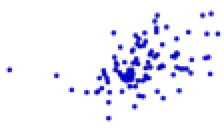
$r = 0.75$



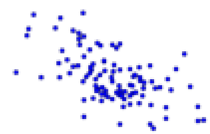
$r = -0.25$



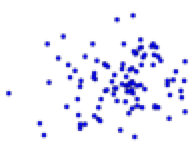
$r = 0.5$



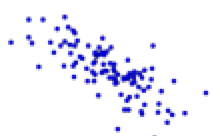
$r = -0.5$



$r = 0.25$



$r = -0.75$



Covariance and Correlation

Range(X) = {0, 1}, Range(Y) = {10, 20}

P(X = 0) = .15 + .3 = .45

P(X = 1) = .3 + .25 = .55

P(Y = 10) = .15 + .3 = .45

P(Y = 20) = .3 + .25 = .55

		Y	
		10	20
	0	.15	.3
X	1	.3	.25

E(X) = 0*.45 + 1*.55 = .55

E(Y) = 10*.45 + 20*.55 = 15.5

E(XY) = 0*10*.15 + 0*20*.3+ 1*10*.3 + 1*20*.25 = 8

cov(X, Y) = E(XY) –E(X)E(Y) = 8 - .55*15.5 = -0.525

E(X^2) = 0*.45 + 1*.55 = .55   E(Y^2) = 100*.45 + 400*.55 = 265

std(X) = sqrt(.55 - .55*.55) =sqrt(.2475) = .4975

std(Y) = sqrt(265 – 15.5*15.5) =sqrt(24.75) = 4.975

corr(X, Y) = cov(X,Y)/std(X)std(Y) = -0.525/.4975*4.975= -.2121

Two-dimensional generative modeling with the bivariate normal distribution

Training set obtained from 130 bottles

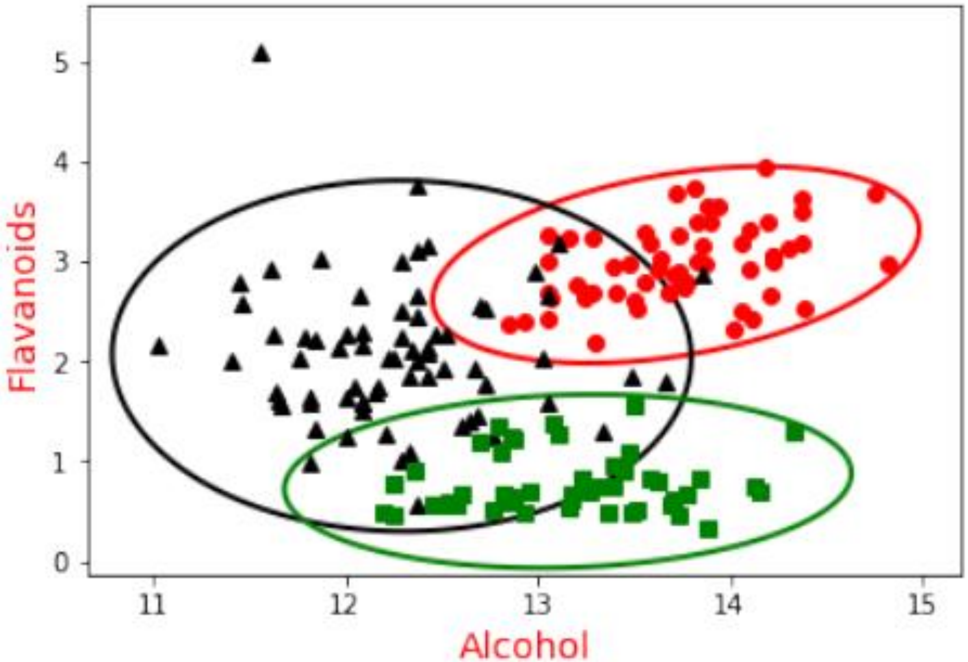
- Winery 1: 43 bottles
- Winery 2: 51 bottles
- Winery 3: 36 bottles
- For each bottle, 13 features:
  'Alcohol', 'Malic acid', 'Ash', 'Alcalinity of ash','Magnesium',
  'Total phenols', 'Flavanoids', 'Nonflavanoid phenols', 'Proanthocyanins',
  'Color intensity', 'Hue', 'OD280/OD315 of diluted wines',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

Also, a separate test set of 48 labeled points.

This time: 'Alcohol' and 'Flavanoids'.

Why more than one feature?

Better **separation** between the classes!



Error rate drops from 29% to 8%.

Bivariate Normal Distribution

Given two variables $x, y \in \mathbb{R}$, the bivariate normal pdf is

$$f(x,y) = \frac{\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right]\right\}}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}$$

where

- $\mu_x \in \mathbb{R}$ and $\mu_y \in \mathbb{R}$ are the marginal means
- $\sigma_x \in \mathbb{R}^+$ and $\sigma_y \in \mathbb{R}^+$ are the marginal standard deviations
- $0 \leq |\rho| < 1$ is the correlation coefficient

X and Y are marginally normal: $X \sim \mathrm{N}(\mu_x, \sigma_x^2)$ and $Y \sim \mathrm{N}(\mu_y, \sigma_y^2)$

Bivariate Normal Distribution

https://en.wikipedia.org/wiki/File:MultivariateNormal.png

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 $\text{std}(X_1)\text{std}(X_2)$.

Bivariate (2-d) Guassian

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 $\left\{ \begin{array}{c} \Sigma_{11} = \mathrm{var}(X_1) \\ \Sigma_{22} = \mathrm{var}(X_2) \\ \Sigma_{12} = \Sigma_{21} = \mathrm{cov}(X_1, X_2) \end{array} \right\}$

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

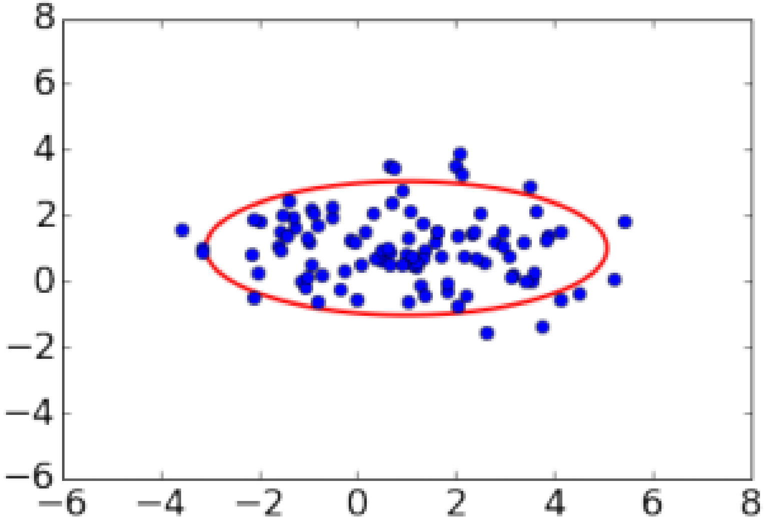
Density of bivariate Guassian

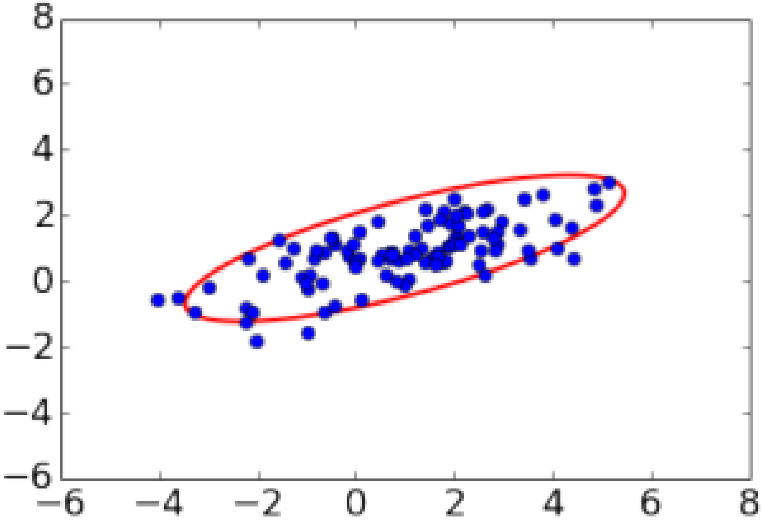
- **Mean** $(\mu_1, \mu_2) \in \mathbb{R}^2$, where $\mu_1 = \mathbb{E}(X_1)$ and $\mu_2 = \mathbb{E}(X_2)$
- **Covariance matrix** $\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$

Density $p(x_1, x_2) = \frac{1}{2\pi|\Sigma|^{1/2}} \exp\left(-\frac{1}{2} \begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix}^T \Sigma^{-1} \begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix} \right)$

Example: Bivariate Guassian

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

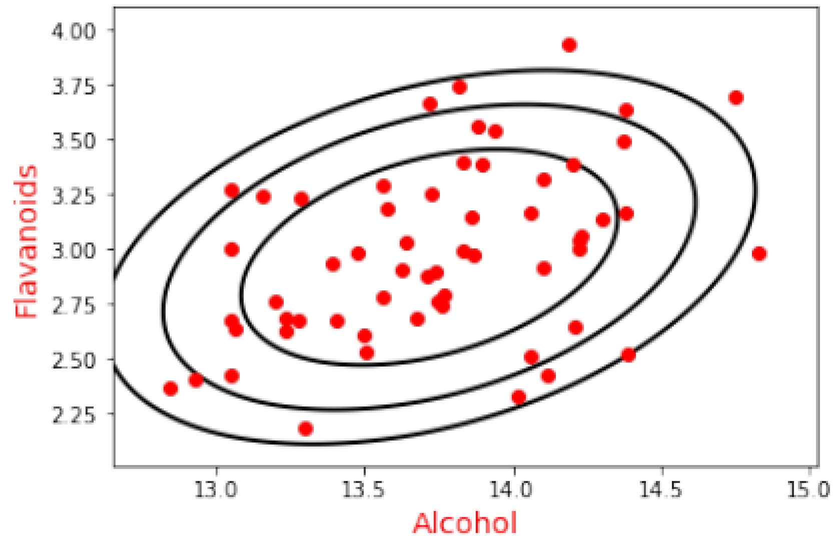




$$\Sigma = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}$$

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

Bivariate Normal Distribution

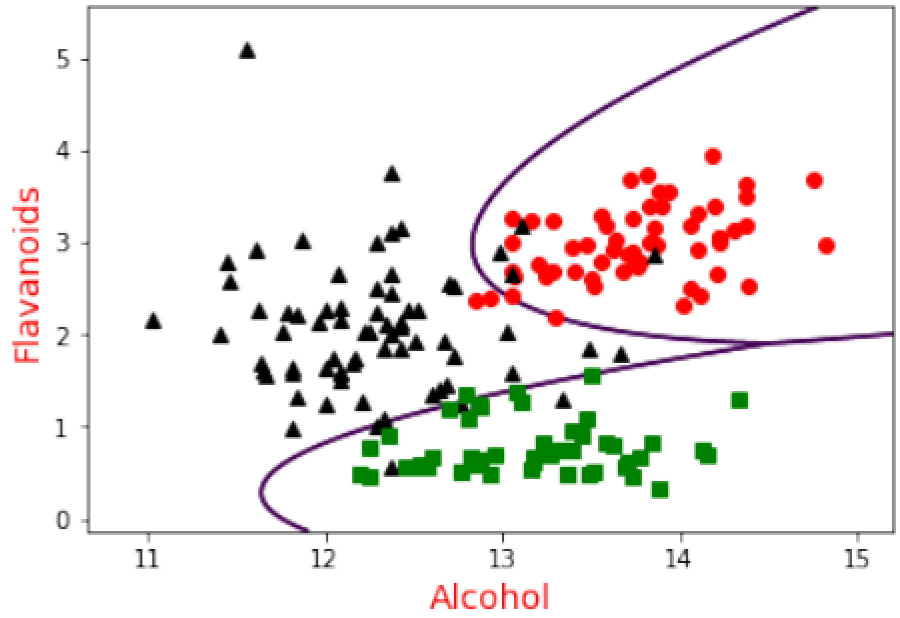


Model class 1 by a bivariate Gaussian, parametrized by:

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

Decision Boundaries

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



Appendix: Links

https://www.youtube.com/channel/UCtYLUTtgS3k1Fg4y5tAhLbw

https://www.khanacademy.org/math/ap-statistics/sampling-distribution-ap/sampling-distribution-proportion/v/normal-conditions-for-sampling-distributions-of-sample-proportions

https://www.statisticshowto.datasciencecentral.com/probability-and-statistics/binomial-theorem/normal-approximation-to-the-binomial/

Demo programs: Python and R

https://machinelearningmastery.com/naive-bayes-classifier-scratch-python/

http://uc-r.github.io/naive_bayes (If you prefer R. Today more Developers/Data Scientists use Python compared to R.)

https://www.analyticsvidhya.com/blog/2017/09/naive-bayes-explained/

https://adataanalyst.com/scikit-learn/linear-classification-method/

Main Point 2

The second step is to determine the probability for an item to be in a class and use it to create the joint distribution function. *Accessing the home of natural law through expansion of awareness makes it possible to bring order and value into any situation, however disorderly it may appear.*

CONNECTING THE PARTS OF KNOWLEDGE WITH THE WHOLENESS OF KNOWLEDGE

1. The Bayes classifier is one of the very best classifiers based on big data.

2. The Bayes classifier is based on Bayes Rule which has its foundation on the probability theory.

---

3. ***Transcendental consciousness*** *is the field beyond all bounds, providing the silent foundation for all computation and structure.*

4. ***Impulses within the Transcendental Field:*** *All the precisely crafted structure of the universe originates from the mistake-free computational dynamics occurring within the transcendent. This unmanifest performance is called by Maharishi Vedic Mathematics.*

5. ***Wholeness moving within itself:*** *In Unity Consciousness, the boundaries and special characteristics that distinguish objects and individuals from one another are appreciated as lively expressions of unified wholeness.*