Bayesian Learning

- Machine Learning by Mitchell-Chp. 6
- Ethem Chp. 3 (Skip 3.6)
- Pattern Recognition & Machine Learning by Bishop Chp. 1

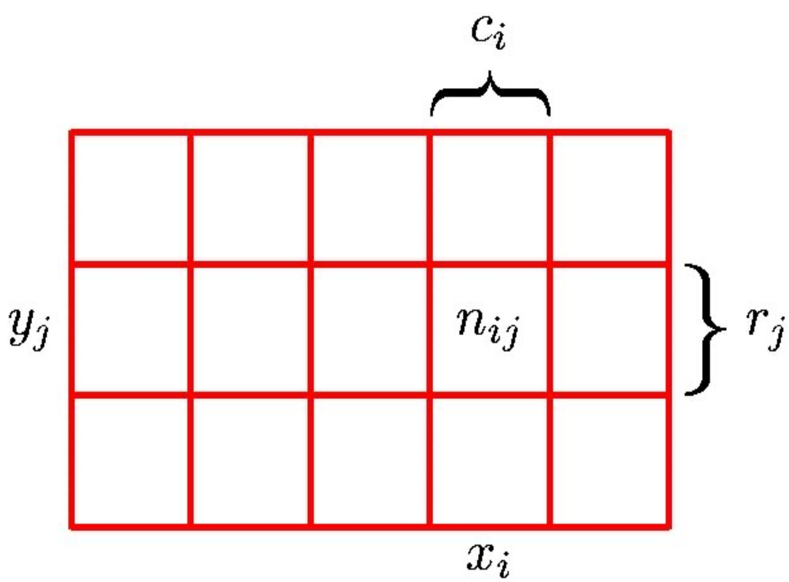
(Pics mostly from the Bishop book)

- Berrin Yanikoglu
  - last edited Oct 2021

Basic Probability

Review

Probability Theory



- **Joint Probability of X and Y**

$$p(X = x_i, Y = y_j) = \frac{n_{ij}}{N}$$

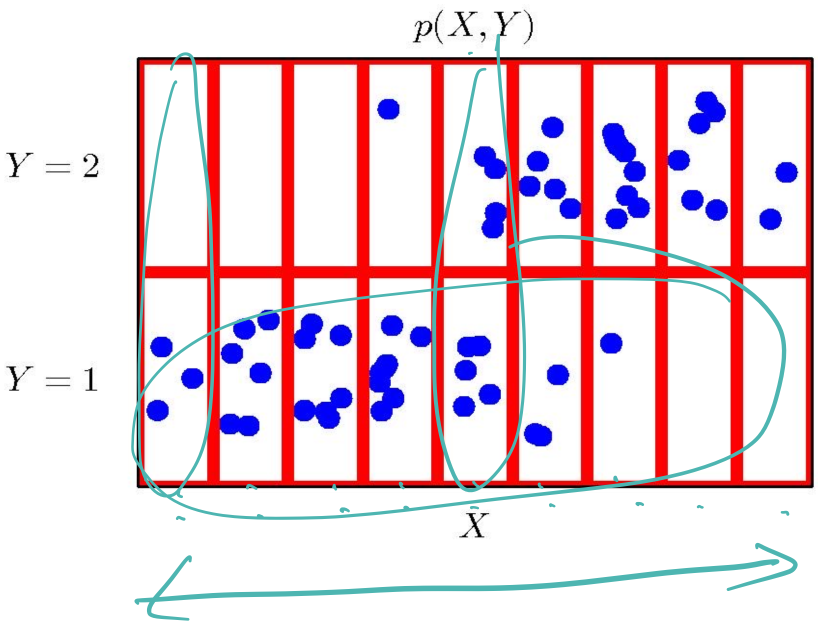
- **Marginal Probability of X**

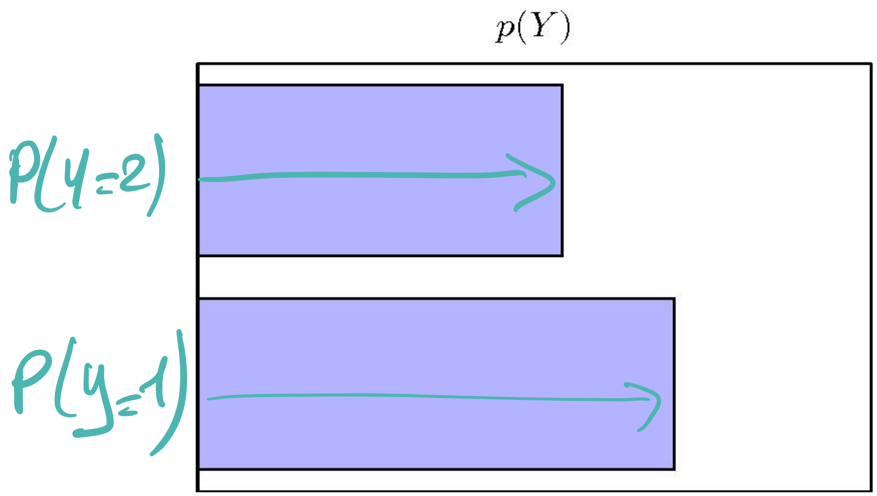
$$p(X = x_i) = \frac{c_i}{N}.$$

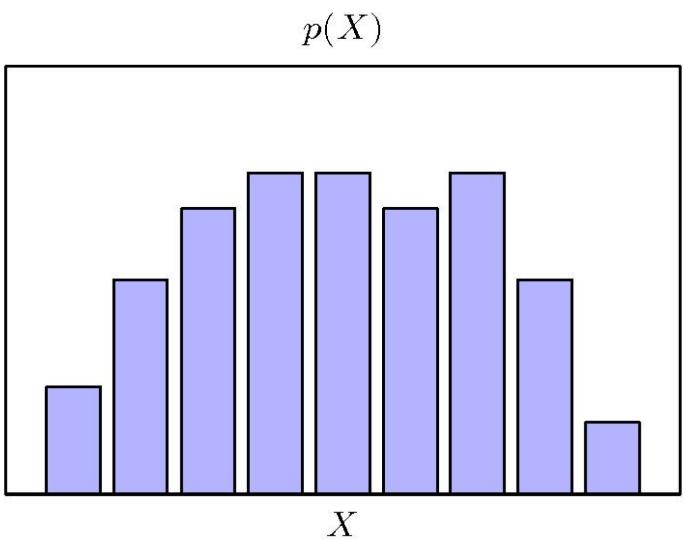
- **Conditional Probability of Y given X**

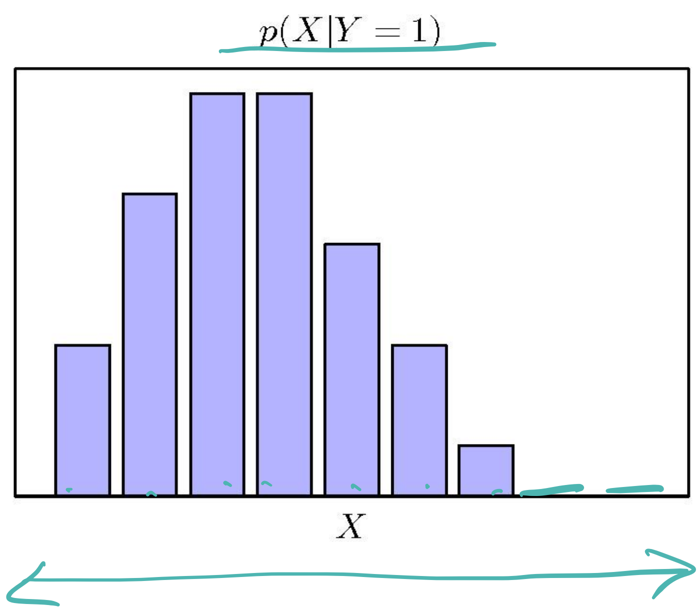
$$p(Y = y_j | X = x_i) = \frac{n_{ij}}{c_i}$$

Probability Theory









Probability Theory

- **Sum Rule**

$$p(X) = \sum_Y p(X, Y)$$

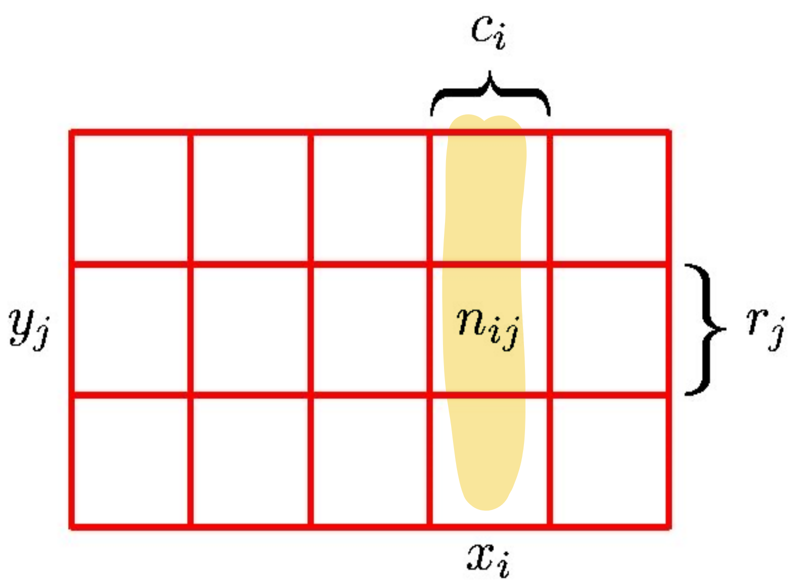
- **Product Rule**

$$p(X, Y) = p(Y|X)p(X)$$

conditional × prior prob.

$$P(x_1 = a, x_2 = b) = P(x_2 = b | x_1 = a) \times P(x_1 = a)$$

Probability Theory



- **Sum Rule**

$$p(X = x_i) = \frac{c_i}{N} = \frac{1}{N} \sum_{j=1}^{L} n_{ij}$$

$$= \sum_{j=1}^{L} p(X = x_i, Y = y_j)$$

Product Rule

$$p(X = x_i, Y = y_j) = \frac{n_{ij}}{N} = \frac{n_{ij}}{c_i} \cdot \frac{c_i}{N}$$

$$= p(Y = y_j | X = x_i) p(X = x_i)$$

Example – In Class

x_2 = height

Weight\Height	Short	Medium	Tall
Low	10	15	5
Medium	8	25	10
Heavy	5	10	12

x_1 = weight

N=100 people with weight and heights given as above.

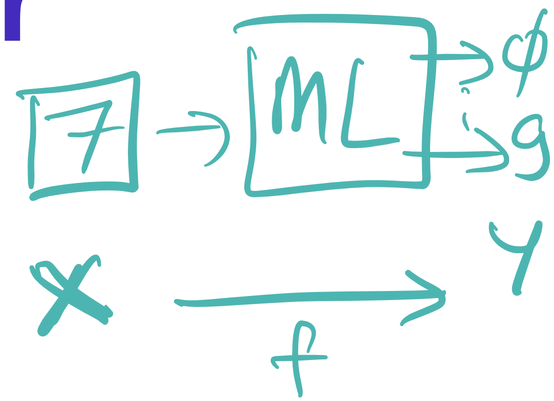
- P (Weight= Low, Height=Tall) = ... 5/100 = 0.05 Joint prob.
- P (Weight= Low | Height=Tall) = ...5/27......... Conditional prob.
- P (Weight= M) = ...43/100 Marginal prob.

(M = Medium)

$$P(x_1 = \text{medium}) = \sum_{v \in \{\text{short}, \text{med.}, \text{tall}\}} P(x_1 = \text{medium}, x_2 = v)$$

Bayesian Decision Theory

Bayes Optimal Classifier



- Goal is to learn f: $\mathbf{X} \longrightarrow Y$
  - $\mathbf{X}$ - features
  - Y - denote the target class

- Suppose you know $P(Y|\mathbf{X})$ exactly, how should you classify?

$$P(y|X)$$

$$P(y=0 \mid \boxed{7}) = 0.1$$

$$P(y=1 \mid \boxed{7}) = 0.3$$

$$\vdots$$

$$P(y=\underline{7} \mid \boxed{7}) = \underline{0.6}$$

$$\vdots$$

$$P(y=9 \mid \boxed{7})$$

Bayes Optimal Classifier

- Goal is to learn f: **X** $\longrightarrow$ Y
  X - features
  Y - denote the target class

- Suppose you know P(Y|**X**) exactly, how should you classify?
  - **Bayes optimal classifier:**

$$Y^* = \underline{\arg\max}_{\underline{\underline{y_k}}} \mathbf{P}\left(Y = \underline{\underline{y_k}} \mid X\right)$$

ϕ - - 9

$Y^* = 7$ (arg. that $\overline{\text{maximizes } P(Y = y_k | X)}$)

Bayesian Decision

- But often, we will not have P(Y | **X**) readily available.
  - Consider diagnosing the problem given BodyAche.
  - Assume it could only be Flu vs Covid19.

- See what you can answer easily?
  - P(Covid19) = 0.3
  - P(Flu) = 0.7
  - P(BodyAche | Flu) = 0.9 // typically easy to estimate
  - P(Flu | BodyAche) = …. // this is of diagnostic interest

- **Bayes Theorem enables us to compute the posterior probabilities** P(Y | **X**) **given priors** P(**X**) **and class conditional probabilities** P(**X** | Y)

Bayes' Theorem

$$p(Y|X) = \frac{p(X|Y)p(Y)}{p(X)}$$

$$p(X) = \sum_Y p(X|Y)p(Y)$$

Product rule

Starting with:
P(C1,X=x) = P(X=x|C_1) P(C_1)

$P(C|X)$

$P(C,X) = P(X|C) \cdot P(C)$
$= P(C|X) \cdot P(X)$

$P(C|X) \cdot P(X) = P(X|C) \cdot P(C) = P(C,X)$

Using this formula for classification problems, we get

$$\boxed{\underline{P(C|X)} = \underline{P(X|C)}\ P(C) / P(X)}$$

what we will use mainly!

posterior probability = α x class conditional probability x prior of the class

$P(X|C) = \frac{P(C|X)\,P(X)}{P\,C}$

$$\boxed{9} \rightarrow \boxed{\text{Bayes Class.}} \rightarrow \begin{matrix} 0 \\ 9 \end{matrix}$$

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)}$$

49.99

10.99

8.99

$$P(C=0) \times P(\boxed{9}|C=0) \rightarrow P(C=0|\boxed{9})$$

$$P(C=9) \times P(\boxed{9}|C=9) \rightarrow P(C=9|\boxed{9})$$

Since $P(X)$ appears for all classes, it can be ignored.

$$P(C|X) = \frac{1}{P(X)} \cdot P(X|C) \times P(C)$$

$$\Downarrow$$

$$P(C|X) \propto P(X|C) \cdot P(C)$$

Bayesian Decision

- You would minimize the number of misclassifications if you choose the class that has the maximum posterior probability:
  - Choose C_1 if $p(C_1|X=x) > p(C_2|X=x)$
  - Choose C_2 otherwise

- Equivalently, since $p(C_1|X=x) = p(X=x|C_1)P(C_1)/P(X=x)$
  - Choose C_1 if $p(X=x|C_1)P(C_1) > p(X=x|C_2)P(C_2)$
  - Choose C_2 otherwise

- Notice that both $p(X=x|C_1)$ and $P(C_1)$ are easier to compute than $P(C_i|x)$.

$$\frac{P(\text{LostTaste} \mid \text{Flu}) \times P(\text{Flu})}{P(\text{''} \mid \text{Covid}) \times P(\text{Covid})}$$

> Bayes Optimal Classifier

$P(\text{Covid} \mid \text{LostTaste})$?

$P(\text{Flu} \mid \text{LostTaste})$

Example

Classify according to height (x)	X <150	X=[150-159]	X=[160-169]	X=[170-179]	X>180
C1=man	10	90	250	300	150
C2=woman	20	200	200	130	50

600 samples in C_2

800 samples in C_1

Total 1400 samples

$$P(C_1, X=x) = \frac{\text{num. samples in corresponding box}}{\text{num. all samples}}$$

//joint probability of C_1 and X

$$P(X=x|C_1) = \frac{\text{num. samples in corresponding box}}{\text{num. of samples in } C_1\text{-row}}$$

//class-conditional probability of X

$$P(C_1) = \frac{\text{num. of of samples in } C_1\text{-row}}{\text{num. all samples}}$$

//prior probability of C_1

Example to Work on (Mitchell book)

Does patient have cancer or not?

A patient takes a lab test and the result comes back positive. The test returns a correct positive result in only 98% of the cases in which the disease is actually present, and a correct negative result in only 97% of the cases in which the disease is not present. Furthermore, .008 of the entire population have this cancer.

$$P(cancer) = 0.008 \rightarrow P(\neg cancer) = 1 - 0.008$$
$$P(+|cancer) = 0.98 \rightarrow P(-|cancer) = 1 - 0.98$$
$$P(+|\neg cancer) = 1 - 0.97 \leftarrow P(-|\neg cancer) = 0.97$$

Someone's test is +. What is the prob. of cancer?

$$P(cancer|+) = P(+|cancer) \cdot P(cancer) \cdot \frac{1}{P(+)}$$

$$P(cancer) = .008, \qquad P(\neg cancer) = .992$$
$$P(\oplus|cancer) = .98, \qquad P(\ominus|cancer) = .02$$
$$P(\oplus|\neg cancer) = .03, \qquad P(\ominus|\neg cancer) = .97$$

Suppose we now observe a new patient for whom the lab test returns a positive result. Should we diagnose the patient as having cancer or not? The maximum a posteriori hypothesis can be found using Equation (6.2):

$$P(\oplus|cancer)P(cancer) = (.98).008 = .0078$$
$$P(\oplus|\neg cancer)P(\neg cancer) = (.03).992 = .0298$$

0.21

0.79

0.0376

- find $\frac{1}{P(+)}$

- or simply normalize to 1.

Thus, $h_{MAP} = \neg cancer$. The exact posterior probabilities can also be determined by normalizing the above quantities so that they sum to 1 (e.g., $P(cancer|\oplus) = \frac{.0078}{.0078+.0298} = .21$). This step is warranted because Bayes theorem states that the

$$P(\neg cancer|+) = P(+|\neg cancer) \cdot P(\neg cancer) \cdot \frac{1}{P(+)}$$

- You should be able:
  - E.g. derive marginal and conditional probabilities given a joint probability table.
  - Use them to compute $P(C_i | x)$ using the Bayes theorem
  - Solve problems that are verbally stated as in the previous slide
  - ...