بسم الله الرحمن الرحیم

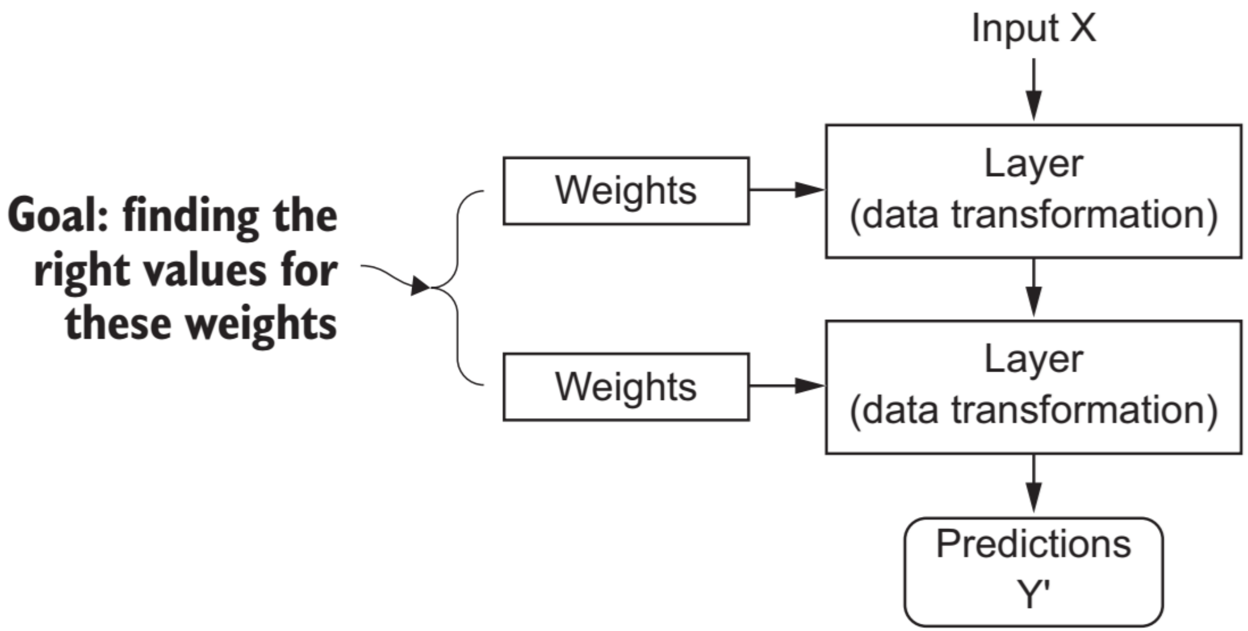
مبانی یادگیری عمیق

مدرس: محمدرضا محمدی

۱۴۰۰

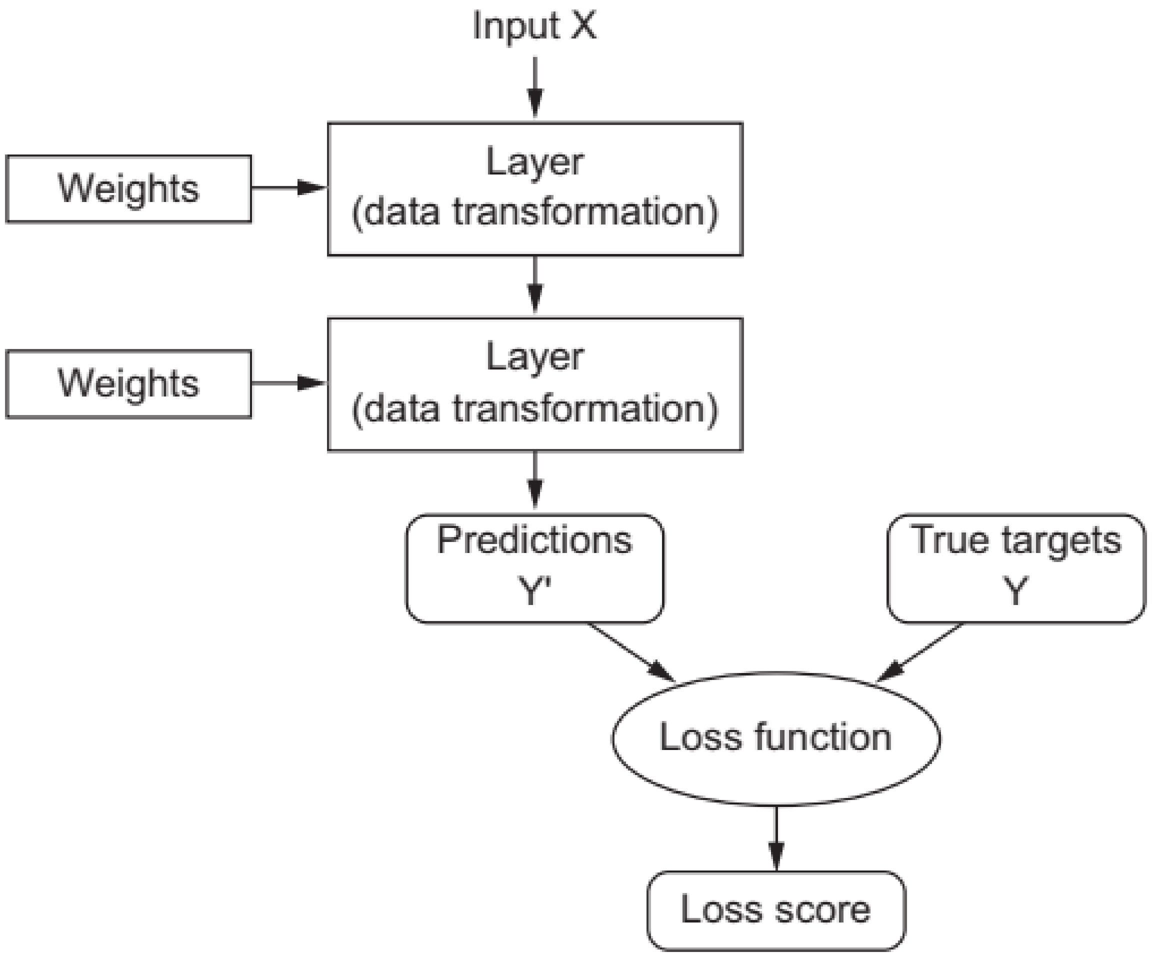
یادگیری عمیق

- یادگیری عمیق به دنبال یافتن یک نگاشت از ورودی‌ها به خروجی‌های مطلوب است
  - این یادگیری از طریق مشاهده نمونه‌های بسیار زیاد از ورودی‌ها و خروجی‌ها حاصل می‌شود
- نیاز است تا وزن‌های تمام لایه‌های شبکه به گونه‌ای تنظیم شوند که بتواند ورودی‌های نمونه را به خروجی‌های مربوطه نگاشت کند

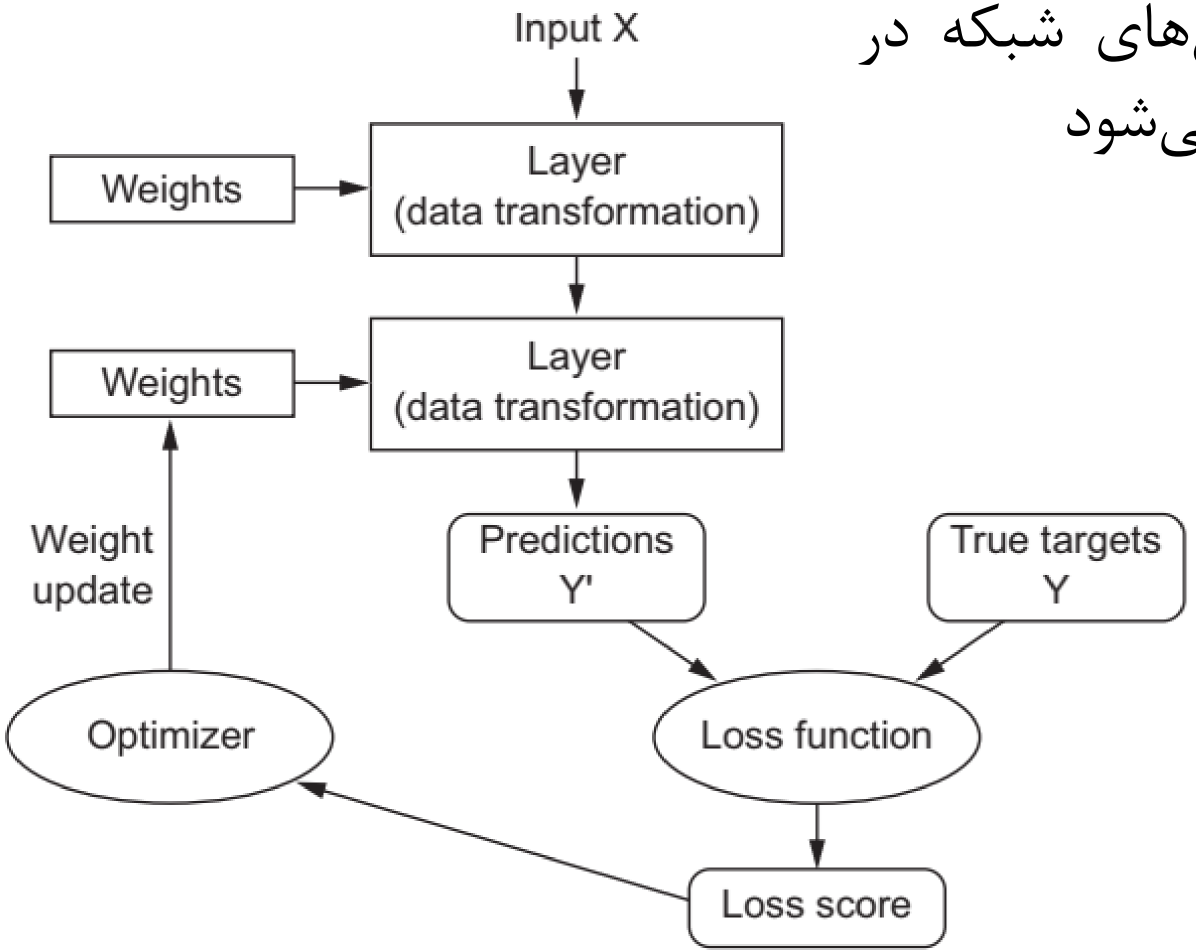


یادگیری عمیق

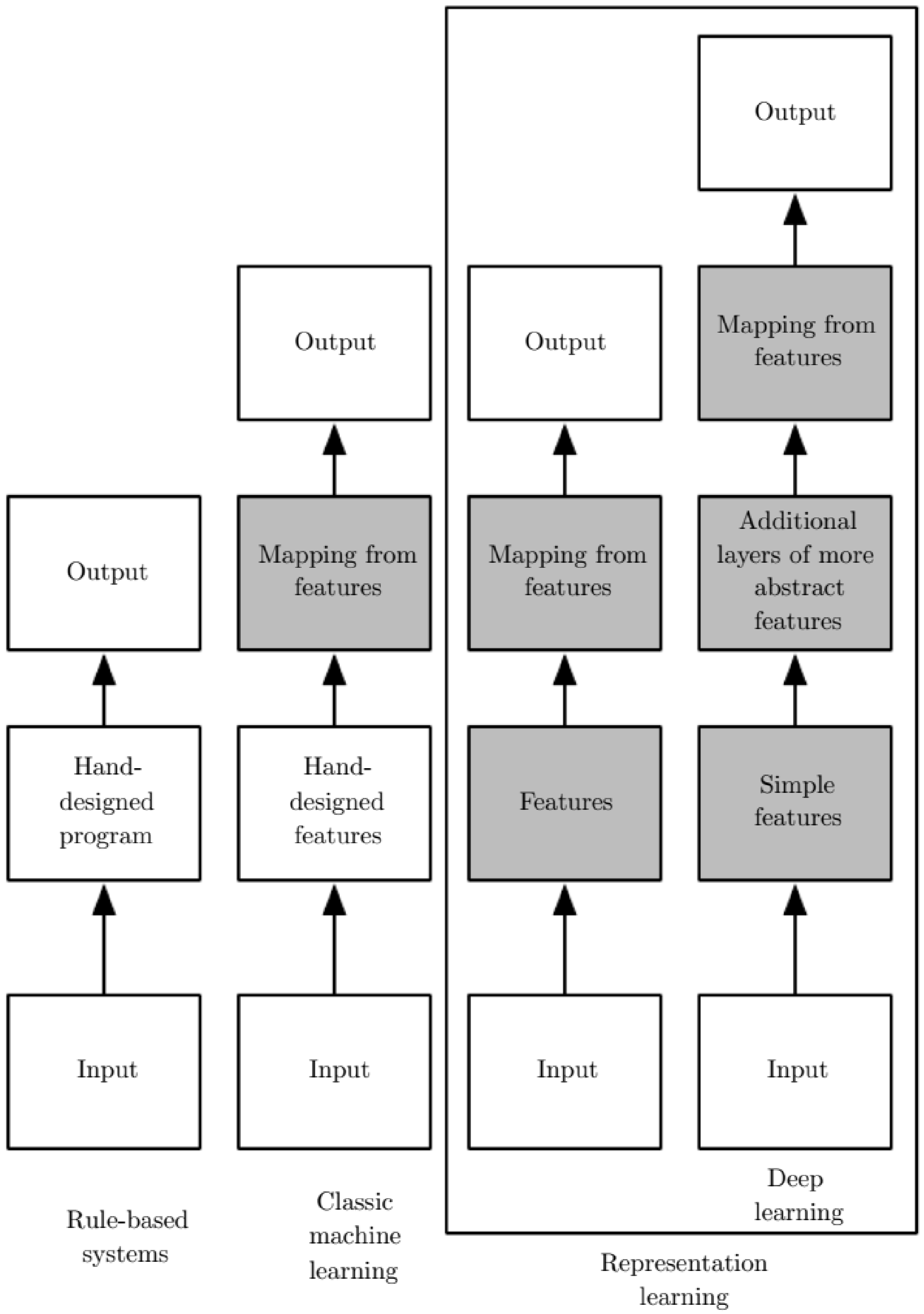
- تابع ضرر (تابع هدف) اندازه‌گیری می‌کند که خروجی‌های شبکه چه مقدار به خروجی‌های مطلوب نزدیک هستند



یادگیری عمیق

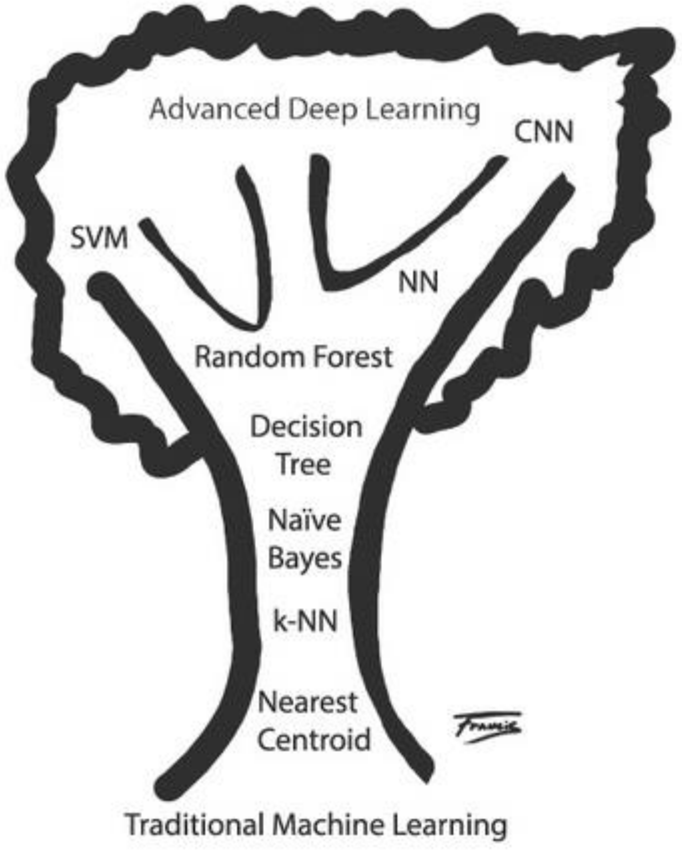


- با استفاده از بازخورد سیگنال تابع ضرر، مقدار وزن‌های شبکه در جهت کاهش تابع ضرر برای نمونه فعلی کمی تنظیم می‌شود
- به یک بهینه‌ساز مناسب نیاز داریم


Output
Output
Output
Mapping from features
Output
Mapping from features
Mapping from features
Additional layers of more abstract features
Hand-designed program
Hand-designed features
Features
Simple features
Input
Input
Input
Input
Rule-based systems
Classic machine learning
Deep learning
Representation learning

پیش از یادگیری عمیق

- یادگیری عمیق همواره بهترین ابزار برای حل مسئله مورد نظر ما نیست
  - ممکن است داده کافی برای آموزش مناسب شبکه‌های عمیق در اختیار نباشد
  - گاهی یک مسئله با الگوریتم‌های دیگر بهتر حل می‌شود
- در ادامه روش مدل‌سازی احتمالی را بررسی می‌کنیم تا مروری بر نظریه احتمالات داشته باشیم



Practical Deep Learning: A Python-Based Introduction
R. Kneusel, 2021.

مدل‌سازی احتمالی

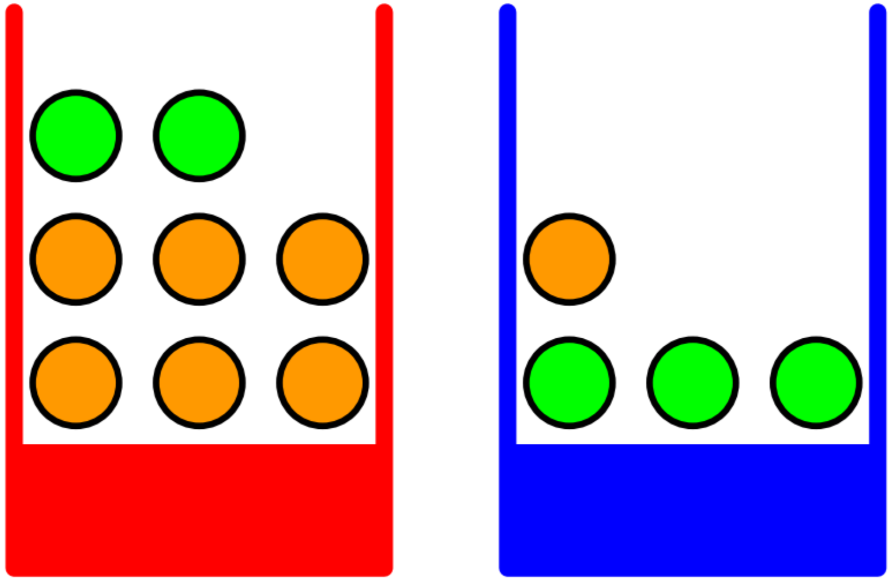
- بخش‌هایی از دانش که ما به آنها دسترسی نداریم، متغیرهای غیرقابل مشاهده نامیده می‌شوند
- در پرتاب سکه، تنها متغیر قابل مشاهده نتیجه پرتاب سکه (شیر یا خط) است
  - آن را به عنوان یک متغیر تصادفی باینری تعریف می‌کنیم
  - فرض کنید $X = 1$ نشان‌دهنده شیر آمدن سکه و $X = 0$ نشان‌دهنده خط آمدن باشد

$$P(X = 1) = p_0$$

$$P(X = 0) = 1 - P(X = 1) = 1 - p_0$$

مدل‌سازی احتمالی

- به عنوان یک مثال ساده، فرض کنید دو جعبه به رنگ‌های آبی و قرمز داریم
- در جعبه آبی ۱ توپ سبز و ۶ توپ نارنجی قرار دارد و در جعبه قرمز ۳ توپ سبز و ۱ توپ نارنجی
- به صورت تصادفی یکی از جعبه‌ها را انتخاب می‌کنیم و از درون آن به صورت تصادفی یک توپ را بر می‌داریم
- جعبه انتخاب شده یک متغیر تصادفی است
- توپ انتخاب شده هم یک متغیر تصادفی است



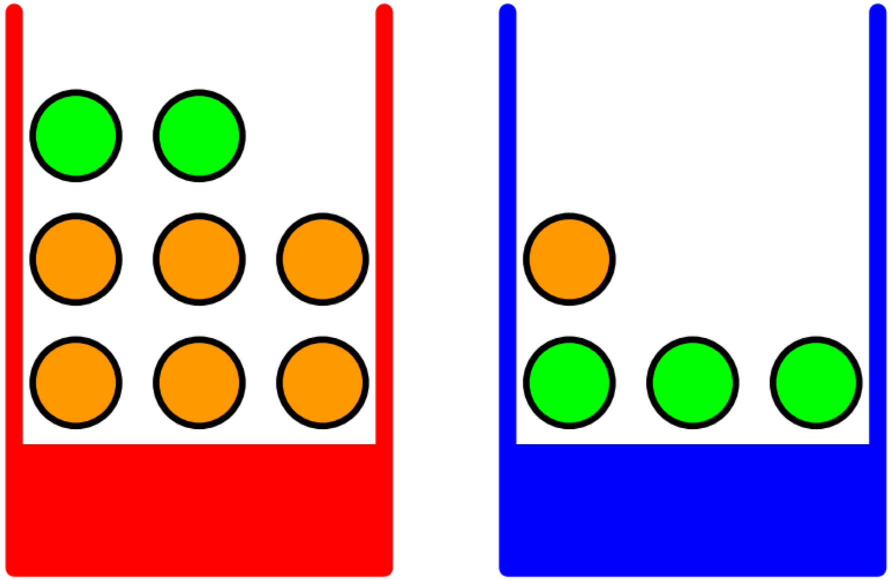
$$Box = \{red, blue\}$$

$$Ball = \{green, orange\}$$

C. Bishop, "*Pattern Recognition and Machine Learning*", 2006

مدل‌سازی احتمالی

- فرض کنید جعبه قرمز با احتمال ۴۰٪ انتخاب می‌شود
- احتمال آنکه یک توپ سبز انتخاب شود چقدر است؟
- اگر بدانیم یک توپ نارنجی انتخاب شده است، احتمال آنکه جعبه انتخاب شده آبی باشد چقدر است؟



$$P(Box = red) = 0.4$$

$$P(Box = blue) = 0.6$$

$$Box = \{red, blue\}$$

$$Ball = \{green, orange\}$$

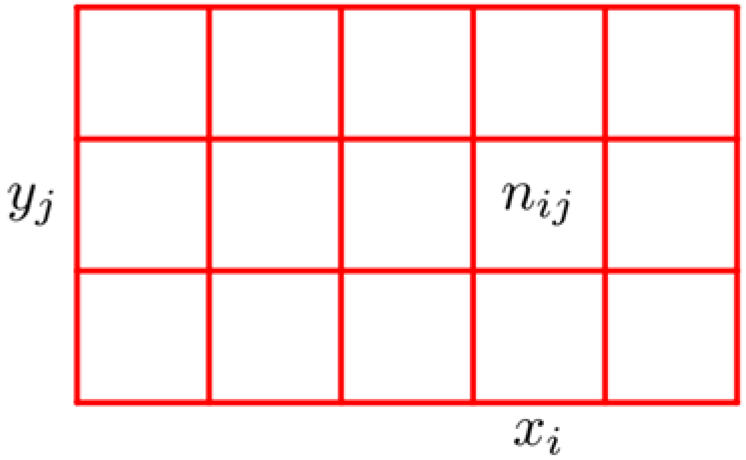
مدل‌سازی احتمالی

- برای بدست آوردن قواعد احتمال، یک مثال که شامل دو متغیر تصادفی X و Y است را در نظر بگیرید
- فرض کنید N مشاهده انجام شده است
- احتمال آنکه متغیر تصادفی X برابر با x_i و Y برابر با y_j باشد، احتمال توام (joint) نامیده می‌شود و برابر است با:

$$X = \{x_i \ \text{for} \ i = 1, 2, \dots, M\}$$

$$Y = \{y_j \ \text{for} \ j = 1, 2, \dots, L\}$$

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

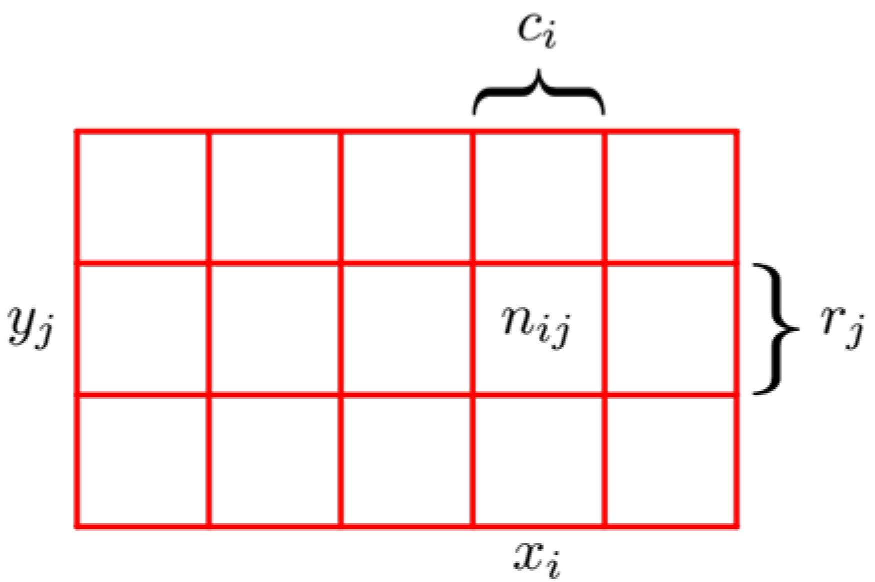


مدل‌سازی احتمالی

- احتمال آنکه مقدار X برابر با x_i باشد مستقل از آنکه Y چه مقداری دارد احتمال حاشیه‌ای (marginal) نامیده می‌شود و برابر است با:

$$P(X = x_i) = \frac{c_i}{N} = \frac{\sum_j n_{ij}}{N} = \sum_{j=1}^{L} P(X = x_i, Y = y_j)$$

- قاعده جمع احتمال



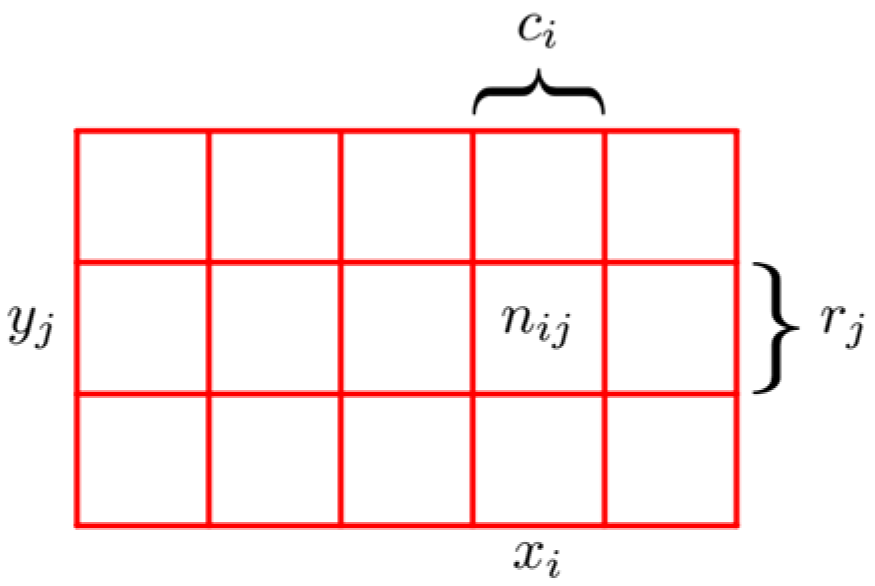
مدل‌سازی احتمالی

- احتمال شرطی (conditional) برای $Y = y_j$ به شرط $X = x_i$:

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

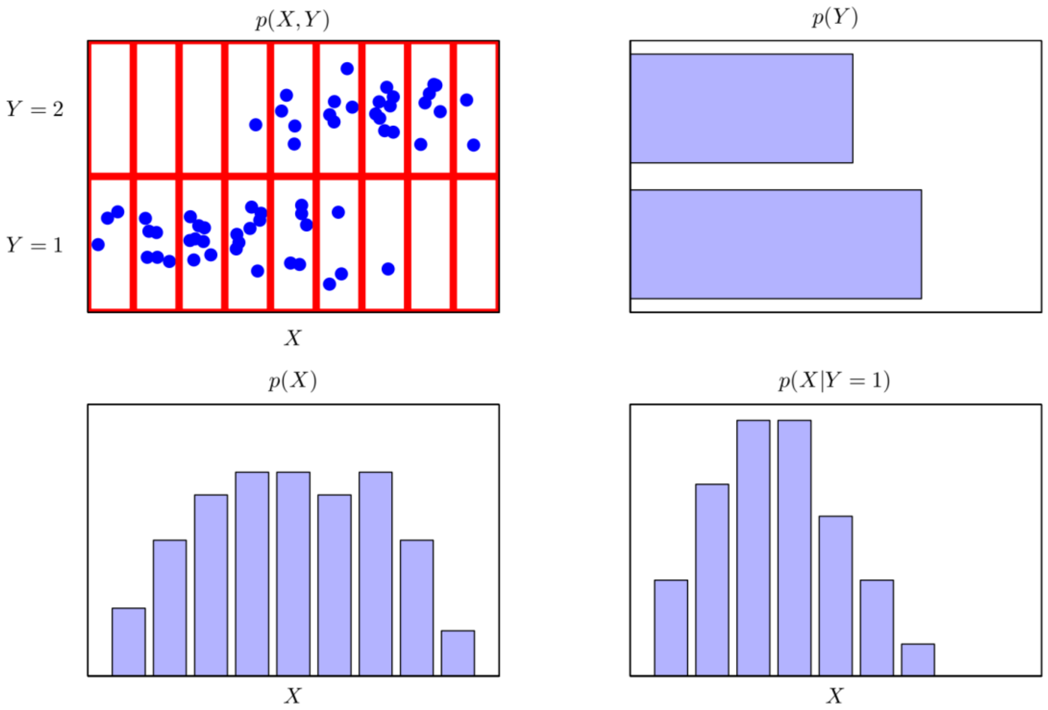
$$P(X = x_i, Y = y_j) = \frac{n_{ij}}{N} = \frac{n_{ij}}{c_i}\frac{c_i}{N}$$

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



- قاعده ضرب احتمال

مدل‌سازی احتمالی



- قاعده جمع:

$$P(X) = \sum_{Y} P(X,Y)$$

- قاعده ضرب:

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

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

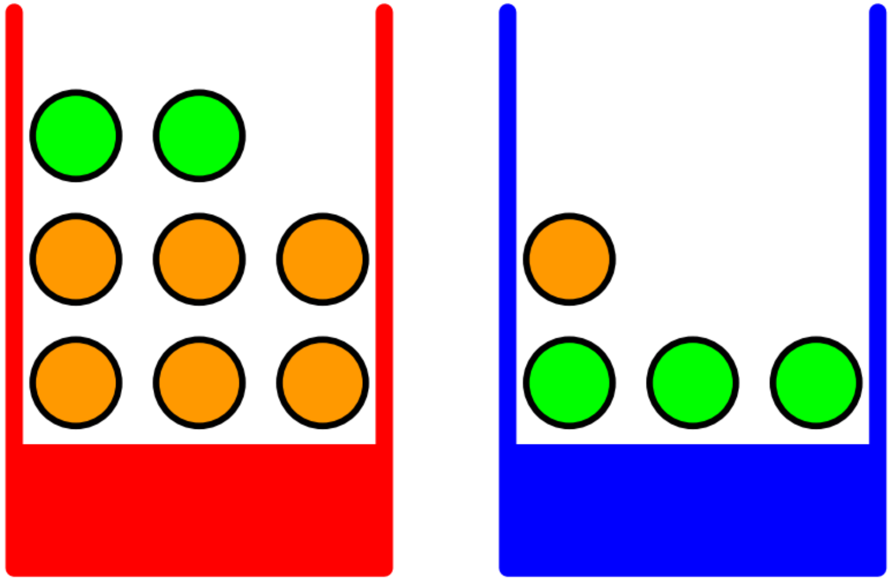
- تئوری Bayes:

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

$$P(X) = \sum_{Y} P(X|Y)P(Y)$$

مدل‌سازی احتمالی

- فرض کنید جعبه قرمز با احتمال ۴۰٪ انتخاب می‌شود
- احتمال آنکه یک توپ سبز انتخاب شود چقدر است؟
- اگر بدانیم یک توپ نارنجی انتخاب شده است، احتمال آنکه جعبه انتخاب شده آبی باشد چقدر است؟



$$Box = \{red, blue\}$$

$$Ball = \{green, orange\}$$

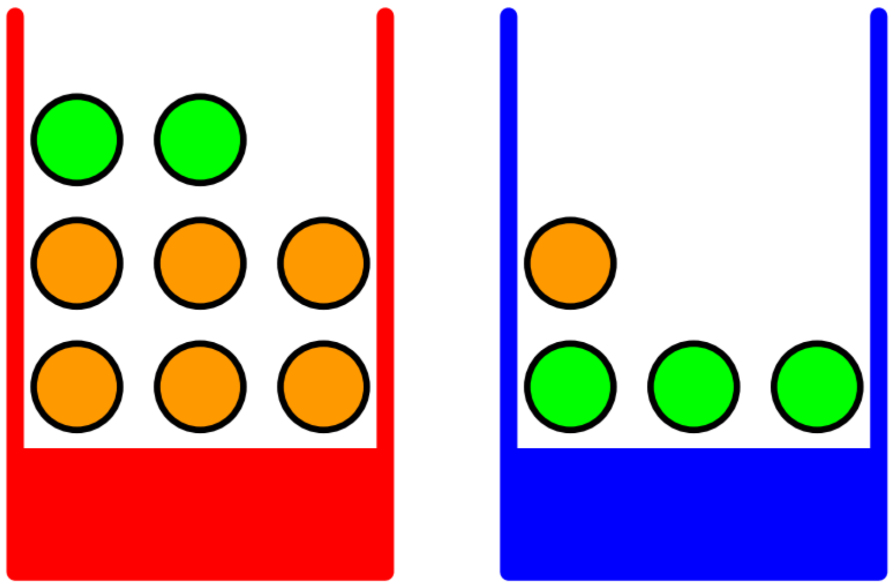
$$P(Box = red) = 0.4$$

$$P(Box = blue) = 0.6$$

مدل‌سازی احتمالی

- احتمال آنکه یک توپ سبز انتخاب شود چقدر است؟

$$P(Ball = green) = P(green|blue)P(blue) + P(green|red)P(red) = \frac{6}{10}\frac{3}{4} + \frac{4}{10}\frac{1}{4} = \frac{11}{20}$$



$$P(Ball = orange) = 9/20$$

$$P(Ball = green|Box = red) = 2/8$$

$$P(Ball = green|Box = blue) = 3/4$$

$$P(Box = red) = 0.4$$

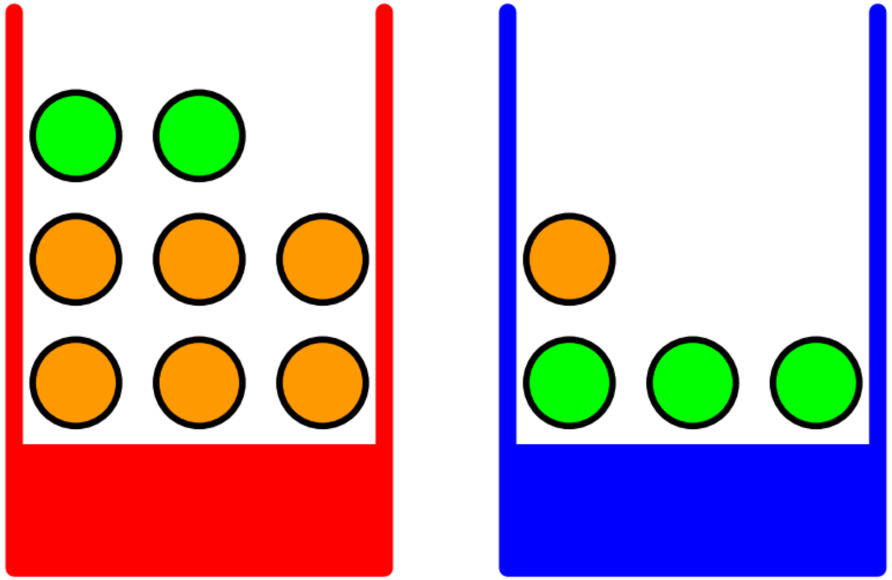
$$P(Box = blue) = 0.6$$

مدل‌سازی احتمالی

- اگر بدانیم یک توپ نارنجی انتخاب شده است، احتمال آنکه جعبه انتخاب شده آبی باشد چقدر است؟

$$P(Box = blue|Ball = orange) = \frac{P(orange|blue)P(blue)}{P(orange)} = \frac{\frac{1}{4}\frac{6}{10}}{\frac{6}{8}\frac{4}{10} + \frac{1}{4}\frac{6}{10}} = \frac{1}{3}$$

$$P(Box = red|Ball = orange) = \frac{2}{3}$$

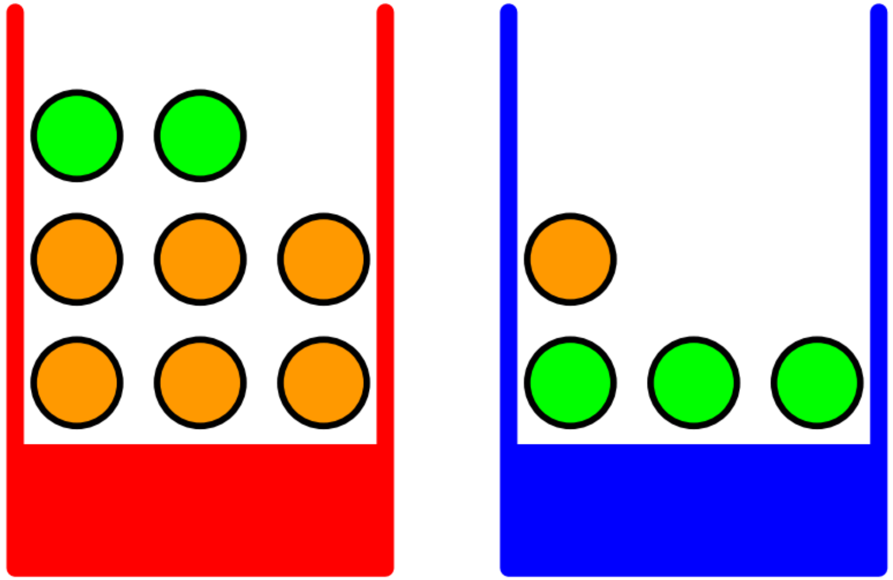


$$P(Box = red) = 0.4$$

$$P(Box = blue) = 0.6$$

مدل‌سازی احتمالی

- چند مفهوم بسیار هم از تئوری Bayes قابل استخراج است:
  - اگر از ما پرسیده شود کدام جعبه انتخاب شده است قبل از آنکه توپ برداشته شده مشخص شود
    - حداکثر اطلاعات ما همان $P(Box)$ است
    - به $P(Box)$ احتمال پیشین (prior) گفته می‌شود
  - اگر بدانیم توپ برداشته شده نارنجی است
    - می‌توانیم با استفاده از تئوری بیز $P(Box|Ball)$ را محاسبه کنیم
    - به $P(Box|Ball)$ احتمال پسین (posterior) گفته می‌شود



$P(Box = red) = 0.4$

$P(Box = blue) = 0.6$

$P(red|orange) = 2/3$

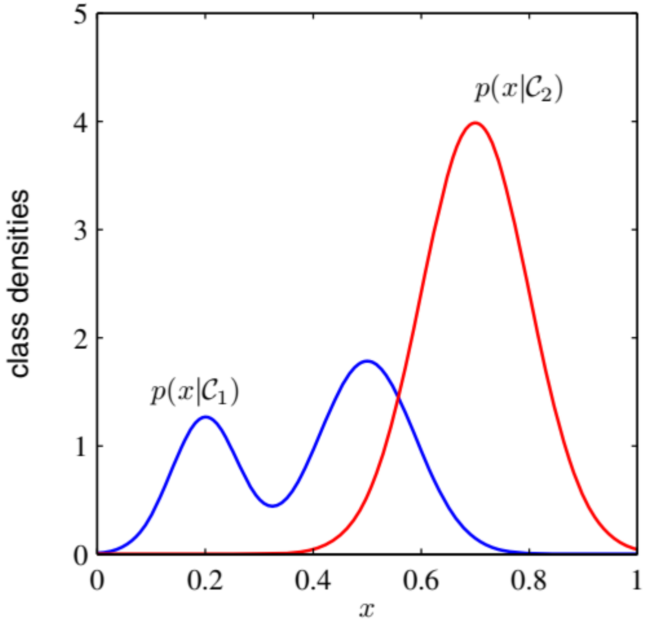
$P(blue|orange) = 1/3$

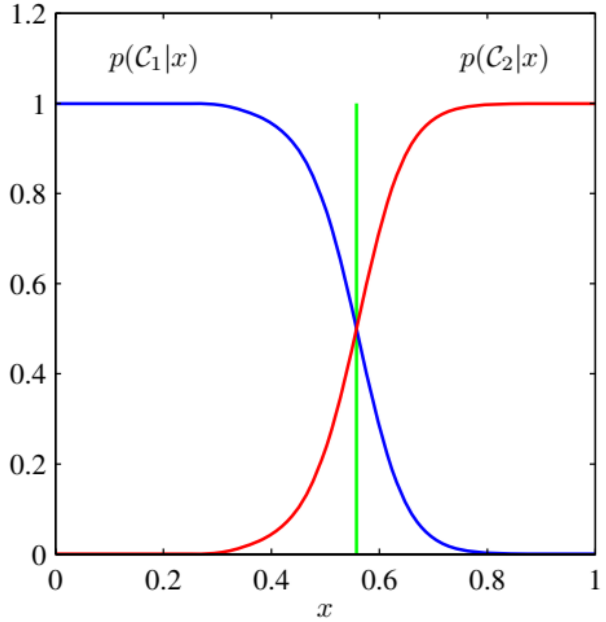
دسته‌بند Bayes

- دسته‌بند Bayes کلاسی را انتخاب می‌کند که بیشترین احتمال پسین را داشته باشد

$$P(C_i|\boldsymbol{x}) = \frac{P(\boldsymbol{x}|C_i)P(C_i)}{P(\boldsymbol{x})} = \frac{P(\boldsymbol{x}|C_i)P(C_i)}{\sum_{k=1}^{K} P(\boldsymbol{x}|C_k)P(C_k)}$$

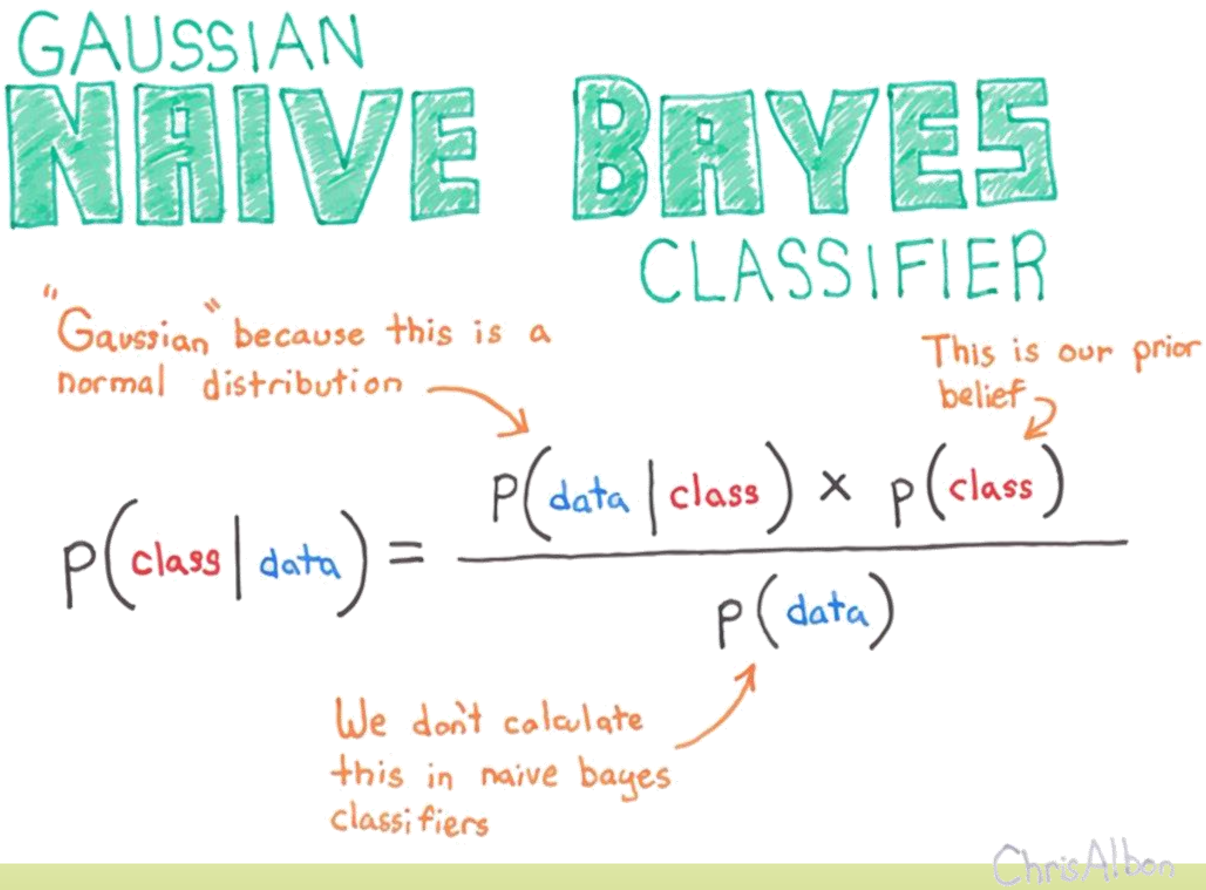
$$\text{choose } C_i \text{ if } P(C_i|\boldsymbol{x}) = \max_k P(C_k|\boldsymbol{x})$$





Naïve Bayes

- یک دسته‌بند مبتنی بر تئوری Bayes است که فرض می‌کند ویژگی‌ها مستقل از هم و دارای توزیع نرمال هستند



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

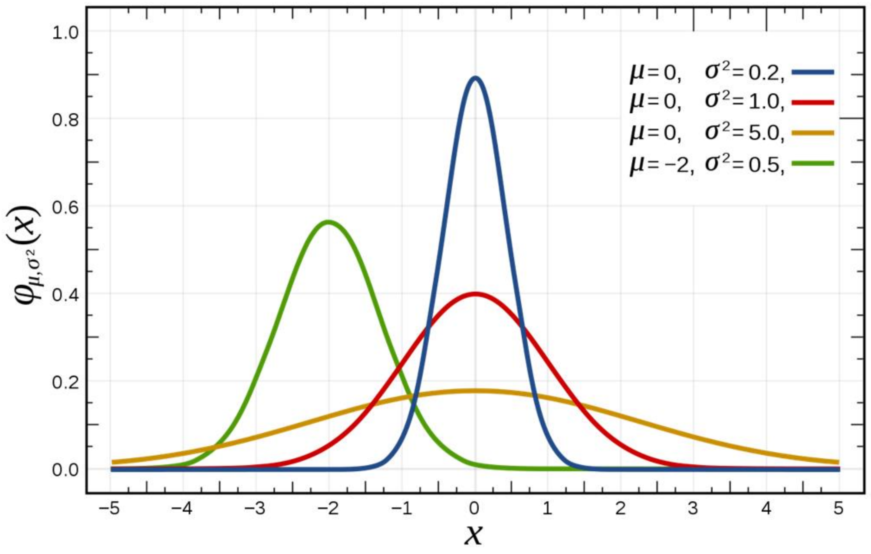
$$P(X) = \prod P(X^k)$$

$$P(X|Y) = \prod P(X^k|Y)$$

مثال

- مسئله: تشخیص مرد یا زن بودن یک انسان
- ویژگی‌ها: قد، وزن و اندازه پا

Person	height (feet)	weight (lbs)	foot size(inches)
male	6	180	12
male	5.92 (5'11")	190	11
male	5.58 (5'7")	170	12
male	5.92 (5'11")	165	10
female	5	100	6
female	5.5 (5'6")	150	8
female	5.42 (5'5")	130	7
female	5.75 (5'9")	150	9

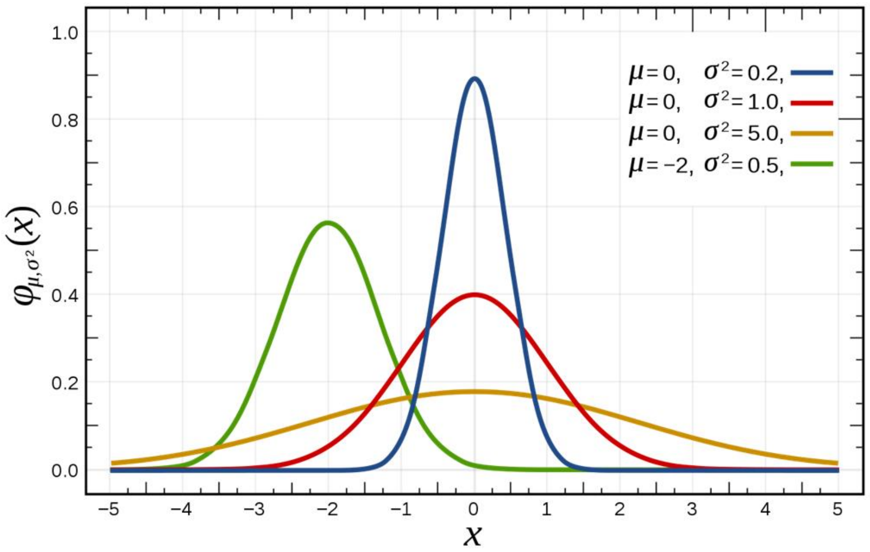


https://en.wikipedia.org/wiki/Naive_Bayes_classifier

مثال

Person	height (feet)	weight (lbs)	foot size(inches)
male	6	180	12
male	5.92 (5'11")	190	11
male	5.58 (5'7")	170	12
male	5.92 (5'11")	165	10
female	5	100	6
female	5.5 (5'6")	150	8
female	5.42 (5'5")	130	7
female	5.75 (5'9")	150	9

Person	height (feet)	weight (lbs)	foot size(inches)
sample	6	130	8



Person	mean (height)	variance (height)	mean (weight)	variance (weight)	mean (foot size)	variance (foot size)
male	5.855	3.5033×10^{-2}	176.25	1.2292×10^{2}	11.25	9.1667×10^{-1}
female	5.4175	9.7225×10^{-2}	132.5	5.5833×10^{2}	7.5	1.6667

مثال

Person	height (feet)	weight (lbs)	foot size(inches)
sample	6	130	8

$$\text{posterior (male)} = \frac{P(\text{male})\, p(\text{height} \mid \text{male})\, p(\text{weight} \mid \text{male})\, p(\text{foot size} \mid \text{male})}{evidence}$$

$$\text{posterior (female)} = \frac{P(\text{female})\, p(\text{height} \mid \text{female})\, p(\text{weight} \mid \text{female})\, p(\text{foot size} \mid \text{female})}{evidence}$$

$$P(\text{male}) = 0.5$$

$$p(\text{height} \mid \text{male}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(6-\mu)^2}{2\sigma^2}\right) \approx 1.5789,$$

Person	mean (height)	variance (height)	mean (weight)	variance (weight)	mean (foot size)	variance (foot size)
male	5.855	3.5033×10^{-2}	176.25	1.2292×10^{2}	11.25	9.1667×10^{-1}
female	5.4175	9.7225×10^{-2}	132.5	5.5833×10^{2}	7.5	1.6667

مثال

Person	height (feet)	weight (lbs)	foot size(inches)
sample	6	130	8

$$\text{posterior (male)} = \frac{P(\text{male})\, p(\text{height} \mid \text{male})\, p(\text{weight} \mid \text{male})\, p(\text{foot size} \mid \text{male})}{evidence}$$

$$\text{posterior (female)} = \frac{P(\text{female})\, p(\text{height} \mid \text{female})\, p(\text{weight} \mid \text{female})\, p(\text{foot size} \mid \text{female})}{evidence}$$

$$P(\text{male}) = 0.5$$

$$p(\text{height} \mid \text{male}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(6-\mu)^2}{2\sigma^2}\right) \approx 1.5789,$$

$$p(\text{weight} \mid \text{male}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(130-\mu)^2}{2\sigma^2}\right) = 5.9881 \cdot 10^{-6}$$

$$p(\text{foot size} \mid \text{male}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(8-\mu)^2}{2\sigma^2}\right) = 1.3112 \cdot 10^{-3}$$

مثال

Person	height (feet)	weight (lbs)	foot size(inches)
sample	6	130	8

$$\text{posterior (male)} = \frac{P(\text{male})\, p(\text{height} \mid \text{male})\, p(\text{weight} \mid \text{male})\, p(\text{foot size} \mid \text{male})}{evidence}$$

$$\text{posterior (female)} = \frac{P(\text{female})\, p(\text{height} \mid \text{female})\, p(\text{weight} \mid \text{female})\, p(\text{foot size} \mid \text{female})}{evidence}$$

$$\text{posterior numerator (male)} = \text{their product} = 6.1984 \cdot 10^{-9}$$

$$\text{posterior numerator (female)} = \text{their product} = 5.3778 \cdot 10^{-4}$$

Person	mean (height)	variance (height)	mean (weight)	variance (weight)	mean (foot size)	variance (foot size)
male	5.855	3.5033×10^{-2}	176.25	1.2292×10^{2}	11.25	9.1667×10^{-1}
female	5.4175	9.7225×10^{-2}	132.5	5.5833×10^{2}	7.5	1.6667