CS189/CS289A
Introduction to Machine Learning
Lecture 5:

Peter Bartlett

February 3, 2015

Outline

Outline

- Two facts from probability theory

Outline

- Two facts from probability theory
- Generative and discriminative models:
  Gaussian class conditionals lead to a logistic posterior.

Outline

- **Two facts from probability theory**
- Generative and discriminative models:
  Gaussian class conditionals lead to a logistic posterior.

Two facts from probability theory

Two facts from probability theory

$$\Pr(A) = \mathbb{E}\Pr(A|X).$$

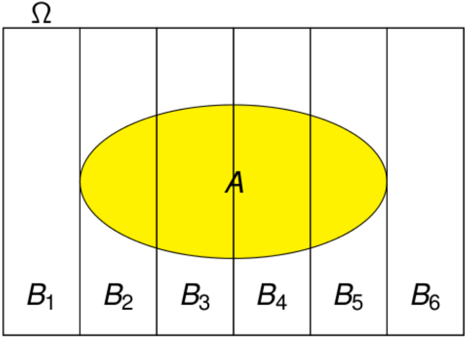
Two facts from probability theory

$$\Pr(A) = \mathbb{E}\Pr(A|X).$$

$$\mathbb{E}(Y) = \mathbb{E}\mathbb{E}(Y|X).$$

Some probability theory

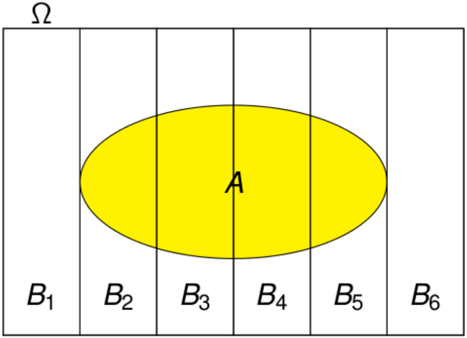
Consider a discrete random variable X. Let $B_i = \{X = i\}$.



Some probability theory

Consider a discrete random variable X. Let $B_i = \{X = i\}$.

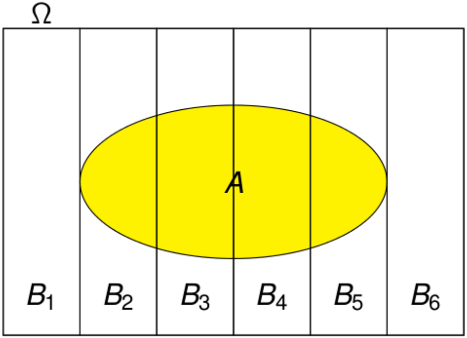
$$\Pr(A) = \sum_i \Pr(A \cap B_i)$$



Some probability theory

Consider a discrete random variable X. Let $B_i = \{X = i\}$.

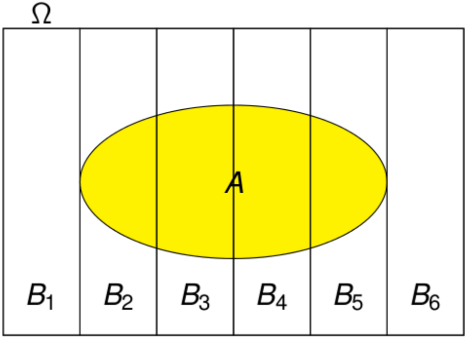
$$\Pr(A) = \sum_i \Pr(A \cap B_i) = \sum_i \Pr(A|B_i)\Pr(B_i)$$



Some probability theory

Consider a discrete random variable X. Let $B_i = \{X = i\}$.

$$\Pr(A) = \sum_i \Pr(A \cap B_i) = \sum_i \Pr(A|B_i)\Pr(B_i) = \mathbb{E}\Pr(A|X).$$



Some probability theory

Example

Some probability theory

Example

- Suppose that, from a set of voters polled,

Some probability theory

Example

- Suppose that, from a set of voters polled,
  - 52% vote Democrat,

Some probability theory

Example

- Suppose that, from a set of voters polled,
  - 52% vote Democrat,
  - 46% vote Republican,

Some probability theory

Example

- Suppose that, from a set of voters polled,
    - 52% vote Democrat,
    - 46% vote Republican,
    - 2% vote Independent.

Example

- Suppose that, from a set of voters polled,
  - 52% vote Democrat,
  - 46% vote Republican,
  - 2% vote Independent.
- Suppose that, of these groups, the proportions who support free community college are:

Some probability theory

Example

- Suppose that, from a set of voters polled,
  - 52% vote Democrat,
  - 46% vote Republican,
  - 2% vote Independent.
- Suppose that, of these groups, the proportions who support free community college are:
  - 86% of Democrat voters,

Some probability theory

Example

- Suppose that, from a set of voters polled,
  - 52% vote Democrat,
  - 46% vote Republican,
  - 2% vote Independent.
- Suppose that, of these groups, the proportions who support free community college are:
  - 86% of Democrat voters,
  - 32% of Republican voters, and

Some probability theory

Example

- Suppose that, from a set of voters polled,
    - 52% vote Democrat,
    - 46% vote Republican,
    - 2% vote Independent.
- Suppose that, of these groups, the proportions who support free community college are:
    - 86% of Democrat voters,
    - 32% of Republican voters, and
    - 48% of Independent voters.

Some probability theory

Example

- Suppose that, from a set of voters polled,
  - 52% vote Democrat,
  - 46% vote Republican,
  - 2% vote Independent.
- Suppose that, of these groups, the proportions who support free community college are:
  - 86% of Democrat voters,
  - 32% of Republican voters, and
  - 48% of Independent voters.
- What proportion of voters support free community college?

Some probability theory

Let F be the event that a random voter supports free community college.
Let P be the party that the random voter votes for.

$$\Pr(F) = \mathbb{E}\Pr(F|P)$$

Some probability theory

Let F be the event that a random voter supports free community college.
Let P be the party that the random voter votes for.

$$\begin{aligned}\Pr(F) &= \mathbb{E}\Pr(F|P) \\ &= \sum_p \Pr(F|P=p)\Pr(P=p)\end{aligned}$$

Some probability theory

Let F be the event that a random voter supports free community college.
Let P be the party that the random voter votes for.

$$\begin{aligned}
\Pr(F) &= \mathbb{E}\Pr(F|P) \\
&= \sum_p \Pr(F|P=p)\Pr(P=p) \\
&= 0.86 \times 0.52 + 0.32 \times 0.46 + 0.48 \times 0.02.
\end{aligned}$$

Some probability theory

Let 1_E denote the random variable

$$1_E = \begin{cases} 1 & \text{when } E \text{ occurs,} \\ 0 & \text{otherwise.} \end{cases}$$

Some probability theory

Let 1_E denote the random variable

$$1_E = \begin{cases} 1 & \text{when } E \text{ occurs,} \\ 0 & \text{otherwise.} \end{cases}$$

Then $\Pr(E) = \mathbb{E}(1_E)$.

Some probability theory

Let 1_E denote the random variable

$$1_E = \begin{cases} 1 & \text{when } E \text{ occurs,} \\ 0 & \text{otherwise.} \end{cases}$$

Then $\Pr(E) = \mathbb{E}(1_E)$.
For any random variable X, we have

$$\Pr(E) = \mathbb{E}\Pr(E|X)$$

Some probability theory

Let 1_E denote the random variable

$$1_E = \begin{cases} 1 & \text{when } E \text{ occurs,} \\ 0 & \text{otherwise.} \end{cases}$$

Then $\Pr(E) = \mathbb{E}(1_E)$.
For any random variable X, we have

$$\begin{aligned} \Pr(E) &= \mathbb{E}\Pr(E|X) \\ \mathbb{E}\,1_E &= \mathbb{E}\left(\mathbb{E}(1_E|X)\right). \end{aligned}$$

Some probability theory

Let 1_E denote the random variable

$$1_E = \begin{cases} 1 & \text{when } E \text{ occurs,} \\ 0 & \text{otherwise.} \end{cases}$$

Then $\Pr(E) = \mathbb{E}(1_E)$.
For any random variable X, we have

$$\begin{aligned} \Pr(E) &= \mathbb{E}\Pr(E|X) \\ \mathbb{E}\,1_E &= \mathbb{E}\left(\mathbb{E}(1_E|X)\right). \end{aligned}$$

This is true not just for indicators of events, but for any (absolutely integrable) random variable Y:

$$\mathbb{E}Y = \mathbb{E}\left(\mathbb{E}(Y|X)\right).$$

Some probability theory

For discrete Y, we can write $Y = \sum_y y 1_{[Y=y]}$

Some probability theory

For discrete Y, we can write $Y = \sum_y y 1_{[Y=y]}$, so

$$\mathbb{E}Y = \mathbb{E}\sum_y y 1_{Y=y}$$

Some probability theory

For discrete Y, we can write $Y = \sum_y y 1_{[Y=y]}$, so

$$\begin{aligned}\mathbb{E}Y &= \mathbb{E}\sum_y y 1_{Y=y} \\ &= \sum_y y \mathbb{E} 1_{Y=y}\end{aligned}$$

Some probability theory

For discrete Y, we can write $Y = \sum_y y 1_{[Y=y]}$, so

$$\begin{aligned} \mathbb{E}Y &= \mathbb{E}\sum_y y 1_{Y=y} \\ &= \sum_y y \mathbb{E} 1_{Y=y} \\ &= \sum_y y \mathbb{E}\mathbb{E}(1_{Y=y}|X) \end{aligned}$$

Some probability theory

For discrete Y, we can write $Y = \sum_y y 1_{[Y=y]}$, so

$$
\begin{aligned}
\mathbb{E}Y &= \mathbb{E}\sum_y y 1_{Y=y} \\
&= \sum_y y \mathbb{E} 1_{Y=y} \\
&= \sum_y y \mathbb{E}\mathbb{E}(1_{Y=y} | X) \\
&= \mathbb{E}\mathbb{E}\left(\sum_y y 1_{Y=y} \middle| X \right)
\end{aligned}
$$

Some probability theory

For discrete Y, we can write $Y = \sum_y y 1_{[Y=y]}$, so

$$\begin{aligned}
\mathbb{E}Y &= \mathbb{E}\sum_y y 1_{Y=y} \\
&= \sum_y y \mathbb{E} 1_{Y=y} \\
&= \sum_y y \mathbb{E}\mathbb{E}(1_{Y=y}|X) \\
&= \mathbb{E}\mathbb{E}\left(\sum_y y 1_{Y=y} \middle| X\right) \\
&= \mathbb{E}\mathbb{E}\left(Y|X\right).
\end{aligned}$$

Some probability theory

Example

Example

- Suppose that, from a set of voters polled,

Some probability theory

Example

- Suppose that, from a set of voters polled,
    - 52% vote Democrat,

Some probability theory

Example

- Suppose that, from a set of voters polled,
  - 52% vote Democrat,
  - 46% vote Republican,

Some probability theory

Example

- Suppose that, from a set of voters polled,
  - 52% vote Democrat,
  - 46% vote Republican,
  - 2% vote Independent.

Some probability theory

Example

- Suppose that, from a set of voters polled,
  - 52% vote Democrat,
  - 46% vote Republican,
  - 2% vote Independent.
- Suppose that average monthly expenditures on spam are:

Some probability theory

Example

- Suppose that, from a set of voters polled,
  - 52% vote Democrat,
  - 46% vote Republican,
  - 2% vote Independent.
- Suppose that average monthly expenditures on spam are:
  - \$86 for Democrat voters,

Some probability theory

Example

- Suppose that, from a set of voters polled,
  - 52% vote Democrat,
  - 46% vote Republican,
  - 2% vote Independent.
- Suppose that average monthly expenditures on spam are:
  - \$86 for Democrat voters,
  - \$32 for Republican voters, and

Some probability theory

Example

- Suppose that, from a set of voters polled,
  - 52% vote Democrat,
  - 46% vote Republican,
  - 2% vote Independent.
- Suppose that average monthly expenditures on spam are:
  - \$86 for Democrat voters,
  - \$32 for Republican voters, and
  - \$48 for Independent voters.

Some probability theory

Example

- Suppose that, from a set of voters polled,
  - 52% vote Democrat,
  - 46% vote Republican,
  - 2% vote Independent.
- Suppose that average monthly expenditures on spam are:
  - $86 for Democrat voters,
  - $32 for Republican voters, and
  - $48 for Independent voters.
- What is the average monthly spam expenditure?

Some probability theory

Example

- Suppose that, from a set of voters polled,
  - 52% vote Democrat,
  - 46% vote Republican,
  - 2% vote Independent.
- Suppose that average monthly expenditures on spam are:
  - \$86 for Democrat voters,
  - \$32 for Republican voters, and
  - \$48 for Independent voters.
- What is the average monthly spam expenditure?

$$\mathbb{E}(S) = \mathbb{E}\mathbb{E}(S|V_P) = 86 \times 0.52 + 32 \times 0.46 + 48 \times 0.02.$$

Outline

- Two facts from probability theory
- **Generative and discriminative models:**
  Gaussian class conditionals lead to a logistic posterior.

Gaussian generative to logistic discriminative models

Class conditionals to posterior

For Gaussian class conditional densities $P(X|Y=+1)$, $P(X|Y=-1)$ (with the same variance), the posterior probability is logistic

Gaussian generative to logistic discriminative models

Class conditionals to posterior

For Gaussian class conditional densities $P(X|Y = +1)$, $P(X|Y = -1)$ (with the same variance), the posterior probability is logistic:

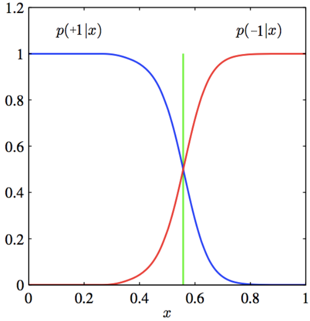
$$P(Y = +1|x) = \frac{1}{1 + \exp\left(-x \cdot \theta - \theta_0\right)}.$$

Gaussian generative to logistic discriminative models

Class conditionals to posterior

For Gaussian class conditional densities $P(X|Y = +1)$, $P(X|Y = -1)$ (with the same variance), the posterior probability is logistic:

$$P(Y = +1|x) = \frac{1}{1 + \exp\left(-x \cdot \theta - \theta_0\right)}.$$



Gaussian generative to logistic discriminative models

- Suppose the class conditional distributions are Gaussian:

$$p(x|y=+1) = \mathcal{N}(\mu_+, \Sigma), \qquad p(x|y=-1) = \mathcal{N}(\mu_-, \Sigma)$$

Gaussian generative to logistic discriminative models

- Suppose the class conditional distributions are Gaussian:

$$p(x|y=+1) = \mathcal{N}(\mu_+, \Sigma), \qquad p(x|y=-1) = \mathcal{N}(\mu_-, \Sigma)$$

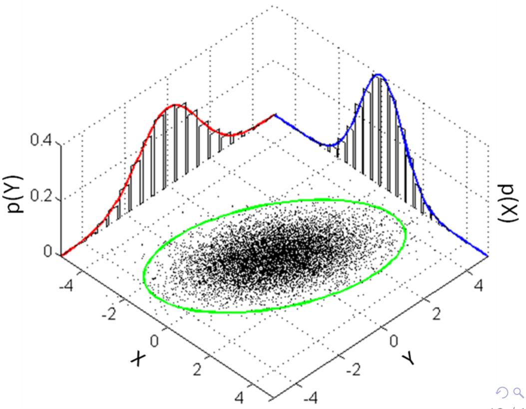
$$p(x|y=+1) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu_+)'\Sigma^{-1}(x-\mu_+)\right).$$

Gaussian generative to logistic discriminative models

- Suppose the class conditional distributions are Gaussian:

$$p(x|y=+1) = \mathcal{N}(\mu_+, \Sigma), \qquad p(x|y=-1) = \mathcal{N}(\mu_-, \Sigma)$$

$$p(x|y=+1) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu_+)'\Sigma^{-1}(x-\mu_+)\right).$$



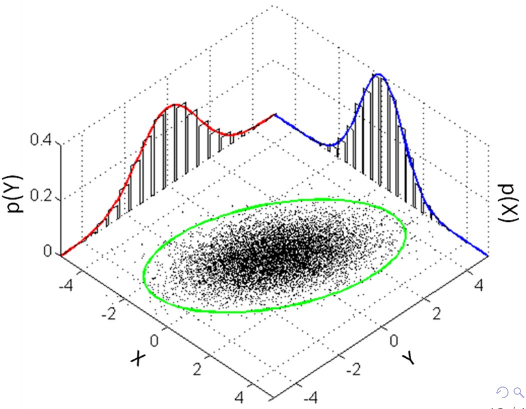
Gaussian generative to logistic discriminative models

- Suppose the class conditional distributions are Gaussian:

$$p(x|y=+1) = \mathcal{N}(\mu_+, \Sigma), \qquad p(x|y=-1) = \mathcal{N}(\mu_-, \Sigma)$$

$$p(x|y=+1) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu_+)'\Sigma^{-1}(x-\mu_+)\right).$$

- Class $+1$ has mean μ_+.
  Class -1 has mean μ_-.



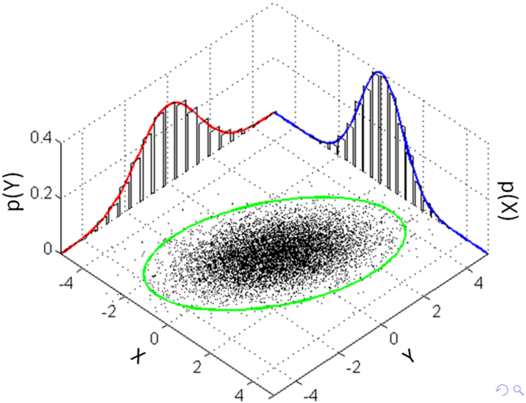
Gaussian generative to logistic discriminative models

- Suppose the class conditional distributions are Gaussian:

$$p(x|y=+1) = \mathcal{N}(\mu_+, \Sigma), \qquad p(x|y=-1) = \mathcal{N}(\mu_-, \Sigma)$$

$$p(x|y=+1) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu_+)'\Sigma^{-1}(x-\mu_+)\right).$$

- Class $+1$ has mean μ_+. Class -1 has mean μ_-.
- Both have covariance matrix Σ.



Gaussian generative to logistic discriminative models

We'll consider the 1-dimensional case, $x, \mu_+, \mu_- \in \mathbb{R}$, $\Sigma = \sigma^2$:

Gaussian generative to logistic discriminative models

We'll consider the 1-dimensional case, $x, \mu_+, \mu_- \in \mathbb{R}$, $\Sigma = \sigma^2$:

$$p(x|y = +1) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x - \mu_+)'\Sigma^{-1}(x - \mu_+)\right)$$

Gaussian generative to logistic discriminative models

We'll consider the 1-dimensional case, $x, \mu_+, \mu_- \in \mathbb{R}$, $\Sigma = \sigma^2$:

$$\begin{aligned} p(x|y=+1) &= \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu_+)'\Sigma^{-1}(x-\mu_+)\right) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu_+)^2}{2\sigma^2}\right). \end{aligned}$$

Gaussian generative to logistic discriminative models

We'll consider the 1-dimensional case, $x, \mu_+, \mu_- \in \mathbb{R}$, $\Sigma = \sigma^2$:

$$p(x|y=+1) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu_+)'\Sigma^{-1}(x-\mu_+)\right)$$
$$= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu_+)^2}{2\sigma^2}\right).$$

Bayes Theorem

Gaussian generative to logistic discriminative models

We'll consider the 1-dimensional case, $x, \mu_+, \mu_- \in \mathbb{R}$, $\Sigma = \sigma^2$:

$$p(x|y = +1) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x - \mu_+)'\Sigma^{-1}(x - \mu_+)\right)$$
$$= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - \mu_+)^2}{2\sigma^2}\right).$$

Bayes Theorem

$$P(Y = +1|x) = \frac{p(x| + 1)P(+1)}{p(x| + 1)P(+1) + p(x| - 1)P(-1)}$$

Gaussian generative to logistic discriminative models

We'll consider the 1-dimensional case, $x, \mu_+, \mu_- \in \mathbb{R}$, $\Sigma = \sigma^2$:

$$
\begin{aligned}
p(x|y=+1) &= \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu_+)'\Sigma^{-1}(x-\mu_+)\right) \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu_+)^2}{2\sigma^2}\right).
\end{aligned}
$$

Bayes Theorem

$$
\begin{aligned}
P(Y=+1|x) &= \frac{p(x|+1)P(+1)}{p(x|+1)P(+1)+p(x|-1)P(-1)} \\
&= \frac{P(+1)\exp\left(-\frac{(x-\mu_+)^2}{2\sigma^2}\right)}{P(+1)\exp\left(-\frac{(x-\mu_+)^2}{2\sigma^2}\right)+P(-1)\exp\left(-\frac{(x-\mu_-)^2}{2\sigma^2}\right)}
\end{aligned}
$$

Gaussian generative to logistic discriminative models

We'll consider the 1-dimensional case, $x, \mu_+, \mu_- \in \mathbb{R}$, $\Sigma = \sigma^2$:

$$p(x|y=+1) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu_+)'\Sigma^{-1}(x-\mu_+)\right)$$
$$= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu_+)^2}{2\sigma^2}\right).$$

Bayes Theorem

$$P(Y=+1|x) = \frac{p(x|+1)P(+1)}{p(x|+1)P(+1) + p(x|-1)P(-1)}$$
$$= \frac{P(+1)\exp\left(-\frac{(x-\mu_+)^2}{2\sigma^2}\right)}{P(+1)\exp\left(-\frac{(x-\mu_+)^2}{2\sigma^2}\right) + P(-1)\exp\left(-\frac{(x-\mu_-)^2}{2\sigma^2}\right)}$$
$$= \frac{1}{1 + \frac{P(-1)}{P(+1)}\exp\left(\frac{(x-\mu_+)^2-(x-\mu_-)^2}{2\sigma^2}\right)}$$

Gaussian generative to logistic discriminative models

Bayes Theorem gives:

$$P(Y = +1|x) = \frac{1}{1 + \frac{P(-1)}{P(+1)} \exp\left(\frac{(x-\mu_+)^2-(x-\mu_-)^2}{2\sigma^2}\right)}$$

Gaussian generative to logistic discriminative models

Bayes Theorem gives:

$$P(Y = +1|x) = \frac{1}{1 + \frac{P(-1)}{P(+1)} \exp\left(\frac{(x-\mu_+)^2 - (x-\mu_-)^2}{2\sigma^2}\right)}$$

$$\frac{P(-1)}{P(+1)} \exp\left(\frac{(x-\mu_+)^2 - (x-\mu_-)^2}{2\sigma^2}\right)$$

Gaussian generative to logistic discriminative models

Bayes Theorem gives:

$$P(Y=+1|x) = \frac{1}{1+\frac{P(-1)}{P(+1)}\exp\left(\frac{(x-\mu_+)^2-(x-\mu_-)^2}{2\sigma^2}\right)}$$

$$\begin{aligned}
&\frac{P(-1)}{P(+1)}\exp\left(\frac{(x-\mu_+)^2-(x-\mu_-)^2}{2\sigma^2}\right) \\
&= \exp\left(\underbrace{\frac{\mu_- - \mu_+}{\sigma^2}} x + \underbrace{\frac{\mu_+^2-\mu_-^2}{2\sigma^2} + \log\left(\frac{P(-1)}{P(+1)}\right)}\right)
\end{aligned}$$

Gaussian generative to logistic discriminative models

Bayes Theorem gives:

$$P(Y = +1|x) = \frac{1}{1 + \frac{P(-1)}{P(+1)} \exp\left(\frac{(x-\mu_+)^2-(x-\mu_-)^2}{2\sigma^2}\right)}$$

$$\begin{aligned}
&\frac{P(-1)}{P(+1)} \exp\left(\frac{(x-\mu_+)^2-(x-\mu_-)^2}{2\sigma^2}\right) \\
&= \exp\left(\underbrace{\frac{\mu_- - \mu_+}{\sigma^2}} x + \underbrace{\frac{\mu_+^2 - \mu_-^2}{2\sigma^2} + \log\left(\frac{P(-1)}{P(+1)}\right)}\right) \\
&= \exp\left(-\theta \cdot x - \theta_0\right).
\end{aligned}$$

Gaussian generative to logistic discriminative models

Bayes Theorem gives:

$$P(Y = +1|x) = \frac{1}{1 + \frac{P(-1)}{P(+1)} \exp\left(\frac{(x-\mu_+)^2-(x-\mu_-)^2}{2\sigma^2}\right)}$$

$$\begin{aligned}
&\frac{P(-1)}{P(+1)} \exp\left(\frac{(x-\mu_+)^2-(x-\mu_-)^2}{2\sigma^2}\right) \\
&= \exp\left(\underbrace{\frac{\mu_- - \mu_+}{\sigma^2}} x + \underbrace{\frac{\mu_+^2 - \mu_-^2}{2\sigma^2} + \log\left(\frac{P(-1)}{P(+1)}\right)}\right) \\
&= \exp\left(-\theta \cdot x - \theta_0\right).
\end{aligned}$$

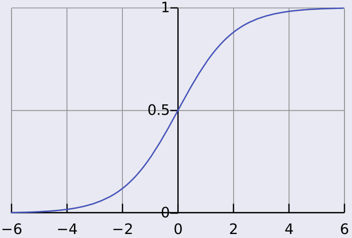
$$P(Y = +1|x) = \frac{1}{1 + \exp\left(-\theta \cdot x - \theta_0\right)}.$$

Gaussian generative to logistic discriminative models

Logistic model: $$P(Y = +1|x) = \frac{1}{1 + \exp\left(-\theta \cdot x - \theta_0\right)}.$$

Gaussian generative to logistic discriminative models

Logistic model: $$P(Y=+1|x) = \frac{1}{1+\exp\left(-\theta \cdot x - \theta_0\right)}.$$

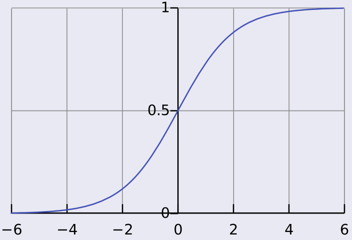


The logistic function $\dfrac{1}{1+e^{-\alpha}}$.

Gaussian generative to logistic discriminative models

Logistic model: $$P(Y=+1|x) = \frac{1}{1+\exp\left(-\theta \cdot x - \theta_0\right)}.$$

$$\theta = \frac{\mu_+ - \mu_-}{\sigma^2}, \qquad \theta_0 = \frac{\mu_-^2 - \mu_+^2}{2\sigma^2} - \log\left(\frac{P(-1)}{P(+1)}\right).$$



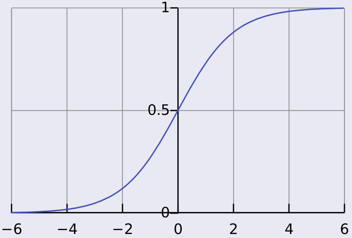
The logistic function $\frac{1}{1+e^{-\alpha}}$.

Gaussian generative to logistic discriminative models

Logistic model: $$P(Y=+1|x) = \frac{1}{1+\exp(-\theta\cdot x - \theta_0)}.$$

$$\theta = \frac{\mu_+ - \mu_-}{\sigma^2}, \qquad \theta_0 = \frac{\mu_-^2 - \mu_+^2}{2\sigma^2} - \log\left(\frac{P(-1)}{P(+1)}\right).$$

- $P(Y=+1|x)$ increases as



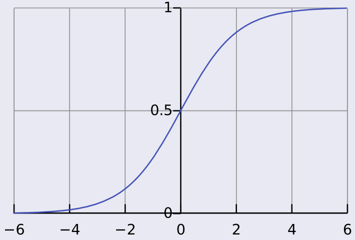
The logistic function $\frac{1}{1+e^{-\alpha}}$.

Gaussian generative to logistic discriminative models

Logistic model: $$P(Y=+1|x) = \frac{1}{1+\exp\left(-\theta\cdot x - \theta_0\right)}.$$

$$\theta = \frac{\mu_+ - \mu_-}{\sigma^2}, \qquad \theta_0 = \frac{\mu_-^2 - \mu_+^2}{2\sigma^2} - \log\left(\frac{P(-1)}{P(+1)}\right).$$

- $P(Y=+1|x)$ increases as
  - x moves from μ_- to μ_+,



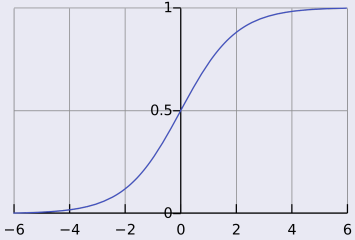
The logistic function $\dfrac{1}{1+e^{-\alpha}}$.

Gaussian generative to logistic discriminative models

Logistic model: $$P(Y=+1|x) = \frac{1}{1+\exp\left(-\theta \cdot x - \theta_0\right)}.$$

$$\theta = \frac{\mu_+ - \mu_-}{\sigma^2}, \qquad \theta_0 = \frac{\mu_-^2 - \mu_+^2}{2\sigma^2} - \log\left(\frac{P(-1)}{P(+1)}\right).$$

- $P(Y=+1|x)$ increases as
  - x moves from μ_- to μ_+,
  - $P(Y=1)$ increases.



The logistic function $\frac{1}{1+e^{-\alpha}}$.

Gaussian generative to logistic discriminative models

- Also true in d dimensions

Gaussian generative to logistic discriminative models

- Also true in d dimensions:

$$\theta = \Sigma^{-1}(\mu_+ - \mu_-), \qquad \theta_0 = \frac{\mu_-' \Sigma^{-1} \mu_- - \mu_+' \Sigma^{-1} \mu_+}{2} - \log \frac{P(-1)}{P(1)}.$$

Gaussian generative to logistic discriminative models

- Also true in d dimensions:

$$\theta = \Sigma^{-1}(\mu_+ - \mu_-), \qquad \theta_0 = \frac{\mu'_-\Sigma^{-1}\mu_- - \mu'_+\Sigma^{-1}\mu_+}{2} - \log\frac{P(-1)}{P(1)}.$$

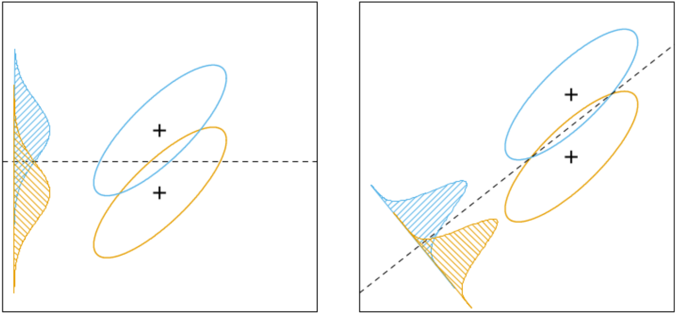


FIGURE 4.9. *Although the line joining the centroids defines the direction of greatest centroid spread, the projected data overlap because of the covariance (left panel). The discriminant direction minimizes this overlap for Gaussian data (right panel).*

Gaussian generative to logistic discriminative models

Gaussian generative to logistic discriminative models

- This is one motivation for logistic regression.

Gaussian generative to logistic discriminative models

- This is one motivation for logistic regression.
  (But logistic regression does not *require* an assumption of Gaussian class conditionals!)

Gaussian generative to logistic discriminative models

- This is one motivation for logistic regression.
  (But logistic regression does not *require* an assumption of Gaussian class conditionals!)
- The posterior $P(Y|X)$ is all we need for classification.

Gaussian generative to logistic discriminative models

- This is one motivation for logistic regression.
  (But logistic regression does not *require* an assumption of Gaussian class conditionals!)
- The posterior $P(Y|X)$ is all we need for classification.
  To go from the posterior $P(Y|X)$ to the class conditionals $P(X|Y)$, we also need to know the marginal $P(X)$.

Outline

- Two facts from probability theory
- Generative and discriminative models:
  Gaussian class conditionals lead to a logistic posterior.