

CS 66: Machine Learning

Prof. Sara Mathieson

Spring 2019



- Office hours **TODAY** 1-3pm
- **Midterm 1 Feb 27** (in lab)
 - Study guide posted
 - You may use a 1-page (front and back) “cheat-sheet”, but no other resources
 - Advice for studying: make cheat sheet, go through notes/slides, redo practice problems, read book for conceptual ideas, make sure study guide makes sense
- **Lab 4 due March 8** (Friday before spring break)
 - Please start soon, but don't neglect midterm studying

Outline for February 22

- Recap Naïve Bayes for discrete features
 - Finish Handout 4
- Examples of Bayes Rule
 - Handout 5
- Naïve Bayes for continuous features

Outline for February 22

- Recap Naïve Bayes for discrete features
 - Finish Handout 4
- Examples of Bayes Rule
 - Handout 5
- Naïve Bayes for continuous features

Handout 4

$$\Theta_1 = \frac{4}{9}, \Theta_2 = \frac{5}{9}$$

f_1	$y=1$	$y=2$
pos	$\frac{2}{5}$	$\frac{2}{3}$
neg	$\frac{3}{5}$	$\frac{1}{3}$

f_2	$y=1$	$y=2$
pos	$\frac{1}{5}$	$\frac{1}{2}$
neg	$\frac{4}{5}$	$\frac{1}{2}$

Laplace

$$\frac{0+1}{3+2}$$

$$|f_2|=2$$

#3 (prediction)

$$\vec{x}_{\text{test}} = [\underbrace{\text{neg}}_{f_1}, \underbrace{\text{pos}}_{f_2}]$$

$$p(y=1 | \vec{x}_{\text{test}}) \propto \Theta_1 \cdot \underbrace{\Theta_{1,\text{neg},1} \cdot \Theta_{2,\text{pos},1}}_{\text{features independent given class label}}$$

$$p(y=k) \prod_{j=1}^P p(x_j | y=k)$$

$$= \frac{4}{9} \cdot \frac{3}{5} \cdot \frac{1}{5} \quad j=1$$

$$= \frac{4}{75}$$

$$p(y=2 | \vec{x}_{\text{test}}) \propto \theta_2 \cdot \theta_{1,\text{neg},2} \cdot \theta_{2,\text{pos},2}$$

$$= \frac{5}{9} \cdot \frac{1}{3} \cdot \frac{1}{2} = \boxed{\frac{5}{54}}$$

$$\boxed{\hat{y} = 2} \star \rightarrow p(y=2 | \vec{x}_{\text{test}}) = \frac{\frac{5}{54}}{\frac{4}{75} + \frac{5}{54}} \approx \boxed{63\%}$$

Informal Quiz: discuss with a partner

- Identify the evidence, prior, posterior, and likelihood in the equation below

$$p(y = k|\mathbf{x}) = \frac{p(y = k)p(\mathbf{x}|y = k)}{p(\mathbf{x})}$$

- What do each of these 4 terms mean?
- Why is the LHS a natural quantity to estimate?
When have we seen it before?
- (open-ended) how could we extend Naïve Bayes to continuous features?

Informal Quiz: discuss with a partner

- Identify the evidence, prior, posterior, and likelihood in the equation below

$$p(y = k|\mathbf{x}) = \frac{p(y = k)p(\mathbf{x}|y = k)}{p(\mathbf{x})}$$

- Evidence:** this is the data (features) we actually observe, which we think will help us predict the outcome we're interested in

Informal Quiz: discuss with a partner

- Identify the evidence, prior, posterior, and likelihood in the equation below

$$p(y = k|\mathbf{x}) = \frac{p(y = k)p(\mathbf{x}|y = k)}{p(\mathbf{x})}$$

- Prior**: without seeing any evidence (data), what is our prior believe about each outcome (intuition: what is the outcome in the population as a whole?)

Informal Quiz: discuss with a partner

- Identify the evidence, prior, posterior, and likelihood in the equation below

$$p(y = k|\mathbf{x}) = \frac{p(y = k)p(\mathbf{x}|y = k)}{p(\mathbf{x})}$$

- Posterior**: this is the quantity we are actually interested in. **Given** the evidence, what is the probability of the outcome?

Informal Quiz: discuss with a partner

- Identify the evidence, prior, posterior, and **likelihood** in the equation below

$$p(y = k|\mathbf{x}) = \frac{p(y = k)p(\mathbf{x}|y = k)}{p(\mathbf{x})}$$

- Likelihood**: given an outcome, what is the probability of observing this set of features?

Informal Quiz: discuss with a partner

- Identify the evidence, prior, posterior, and likelihood in the equation below

$$p(y = k|\mathbf{x}) = \frac{p(y = k)p(\mathbf{x}|y = k)}{p(\mathbf{x})}$$

- Why is the LHS a natural quantity to estimate?
When have we seen it before?

Informal Quiz: discuss with a partner

- Identify the evidence, prior, posterior, and likelihood in the equation below

$$p(y = k|\mathbf{x}) = \frac{p(y = k)p(\mathbf{x}|y = k)}{p(\mathbf{x})}$$

- Why is the LHS a natural quantity to estimate?
When have we seen it before?

We are given a new data point and want to classify/predict what class it belongs to. Probability is important for uncertainty!
Seen before: logistic regression, even k-nearest neighbors

Outline for February 22

- Recap Naïve Bayes for discrete features
 - Finish Handout 4
- Examples of Bayes Rule
 - Handout 5
- Naïve Bayes for continuous features

Example (email)

$$P(\text{spam} | \text{words}) = \frac{P(\text{spam})P(\text{words} | \text{spam})}{P(\text{words})}$$

$$P(\text{words}) = \sum_{\text{vals in spam status}} P(\text{spam status}, \text{words})$$

$$= P(\text{spam}, \text{words}) + P(\overline{\text{spam}}, \text{words})$$

$$\rightarrow = P(\text{spam}, \text{words}) + P(\overline{\text{spam}}, \text{words})$$

spam)

*

status, words)

s) + P($\overline{\text{spam}}$, words)

$$\rightarrow = \left. \begin{aligned} &P(\text{spam})P(\text{words}|\text{spam}) \\ &+ P(\overline{\text{spam}})P(\text{words}|\overline{\text{spam}}) \end{aligned} \right\} \underline{\text{normalizer}}$$

$$\boxed{P(\text{spam}|\text{words}) + P(\overline{\text{spam}}|\text{words}) = 1}$$

Handout 5

$$\left. \begin{aligned} P(\text{disease}) &= P(D) = \frac{1}{100} \\ P(\text{healthy}) &= P(H) = \frac{99}{100} \end{aligned} \right\} \text{prior}$$

test

test with 90% accuracy.

$$P(\text{pos} | D) = \frac{9}{10}$$

$$P(\text{neg} | H) = \frac{9}{10}$$

$$\underline{\underline{Q: P(D|pos) = \frac{p(D)p(pos|D)}{p(pos)}}} \quad \swarrow \text{expand}$$

$$= \frac{p(D)p(pos|D)}{p(D)p(pos|D) + p(H)p(pos|H)}$$

$$= \frac{\frac{1}{100} \cdot \frac{9}{10}}{\frac{1}{100} \cdot \frac{9}{10} + \frac{99}{100} \cdot \frac{1}{10}}$$

$$= \frac{9}{9 + 99} = \frac{9}{108} = \boxed{\frac{1}{12}} \approx 8.25\%$$

Outline for February 22

- Recap Naïve Bayes for discrete features
 - Finish Handout 4
- Examples of Bayes Rule
 - Handout 5
- Naïve Bayes for continuous features

Informal Quiz: discuss with a partner

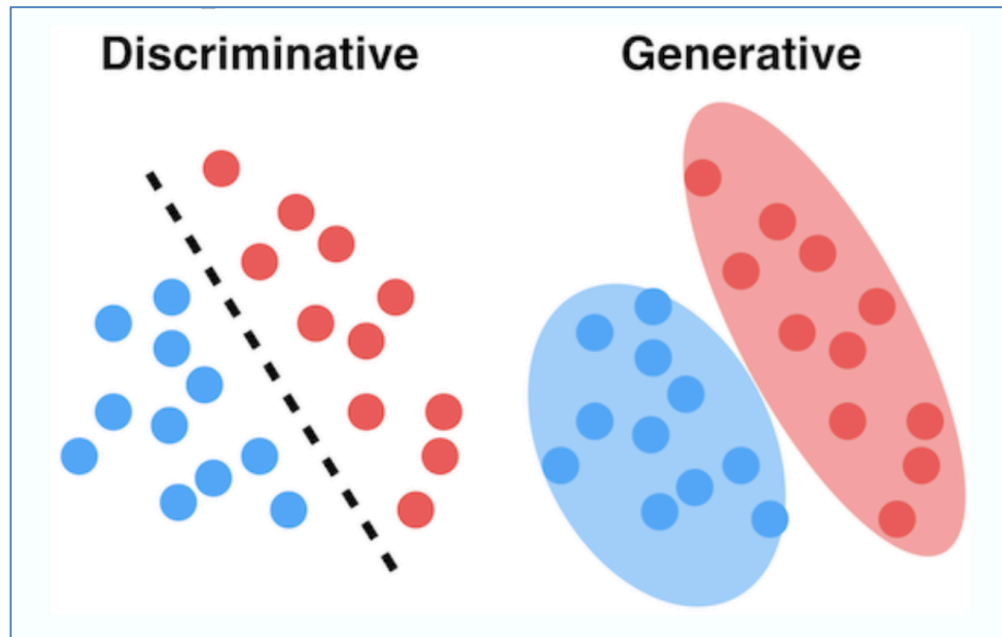
- Identify the evidence, prior, posterior, and likelihood in the equation below

$$p(y = k|\mathbf{x}) = \frac{p(y = k)p(\mathbf{x}|y = k)}{p(\mathbf{x})}$$

- (open-ended) how could we extend Naïve Bayes to continuous features?

Discriminative vs. Generative

- Regression: discriminative model → finds decision boundary
- Naïve Bayes: generative model → estimates probability distributions



Example with $K=2$, $p=1$

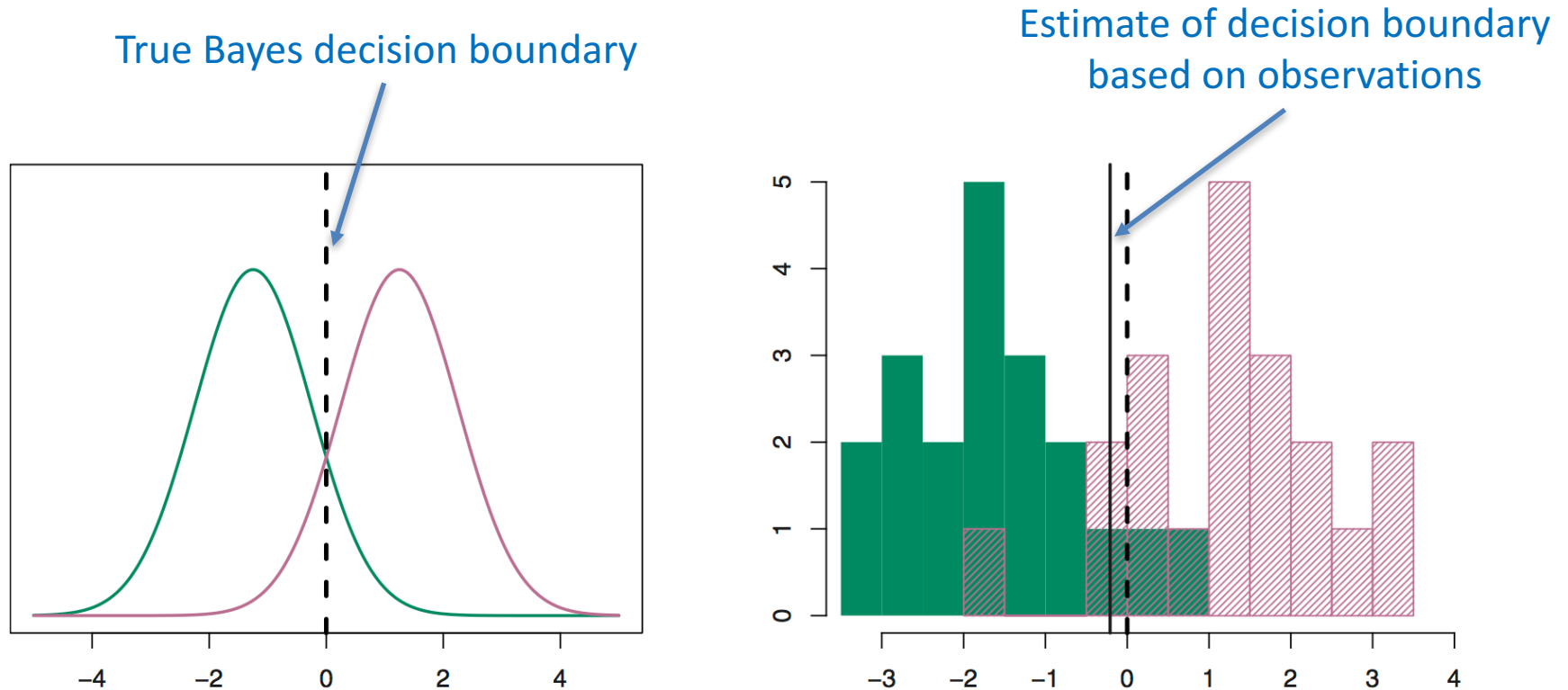


FIGURE 4.4. Left: Two one-dimensional normal density functions are shown. The dashed vertical line represents the Bayes decision boundary. Right: 20 observations were drawn from each of the two classes, and are shown as histograms.

Example with $K=3$, $p=2$

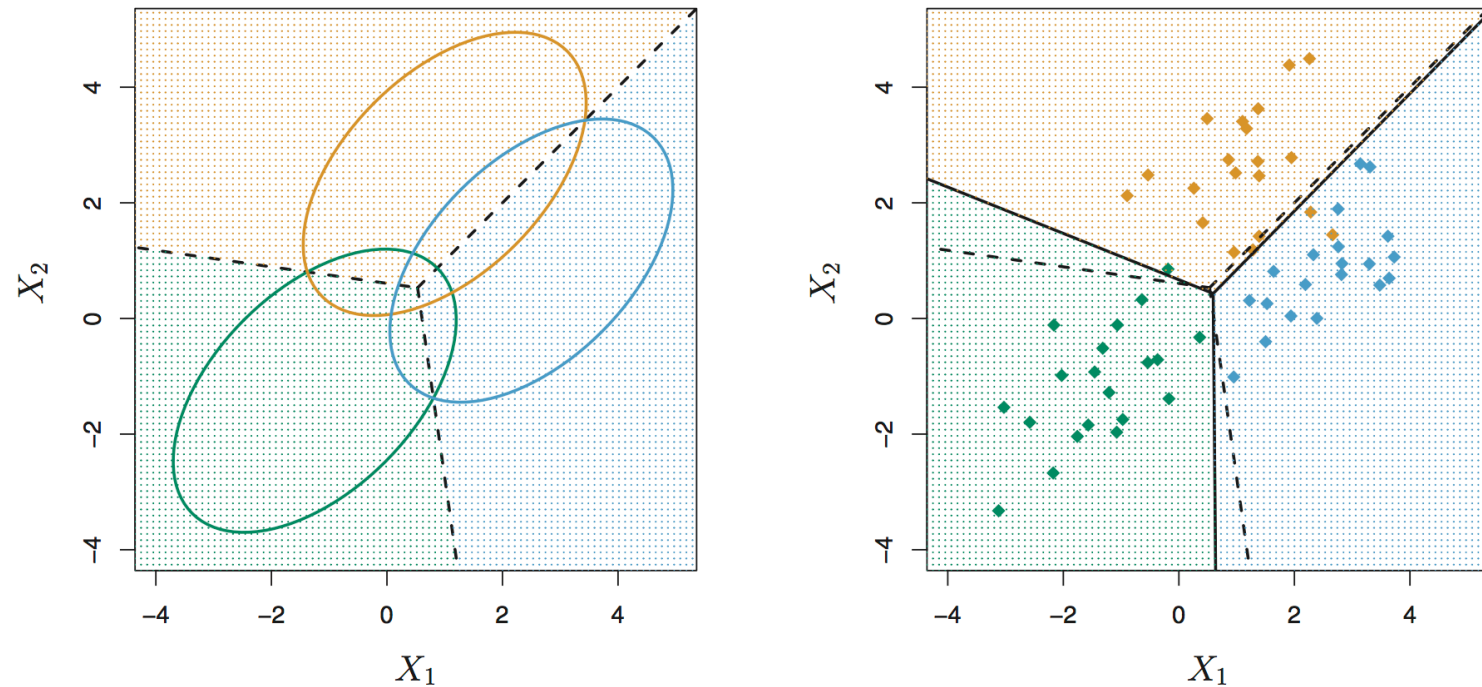


FIGURE 4.6. An example with three classes. The observations from each class are drawn from a multivariate Gaussian distribution with $p = 2$, with a class-specific mean vector and a common covariance matrix. Left: Ellipses that contain 95 % of the probability for each of the three classes are shown. The dashed lines are the Bayes decision boundaries. Right: 20 observations were generated from each class, and the corresponding LDA decision boundaries are indicated using solid black lines. The Bayes decision boundaries are once again shown as dashed lines.