Generative learning algorithms

Kookjin Lee

(kookjin.Lee@asu.edu)

The contents of this course, including lectures and other instructional materials, are copyrighted materials. Students may not share outside the class, including uploading, selling or distributing course content or notes taken during the conduct of the course. Any recording of class sessions is authorized only for the use of students enrolled in this course during their enrollment in this course. Recordings and excerpts of recordings may not be distributed to others. (see ACD 304 – 06 , “ Commercial Note Taking Services ” and ABOR Policy 5 - 308 F.14 for more information).

Discriminant learning algorithms and generative learning algorithm

- So far, we build a model

$$p(y|x)$$

- Generative learning algorithm

$$p(x|y)$$

Bayes rule

- The posterior distribution:

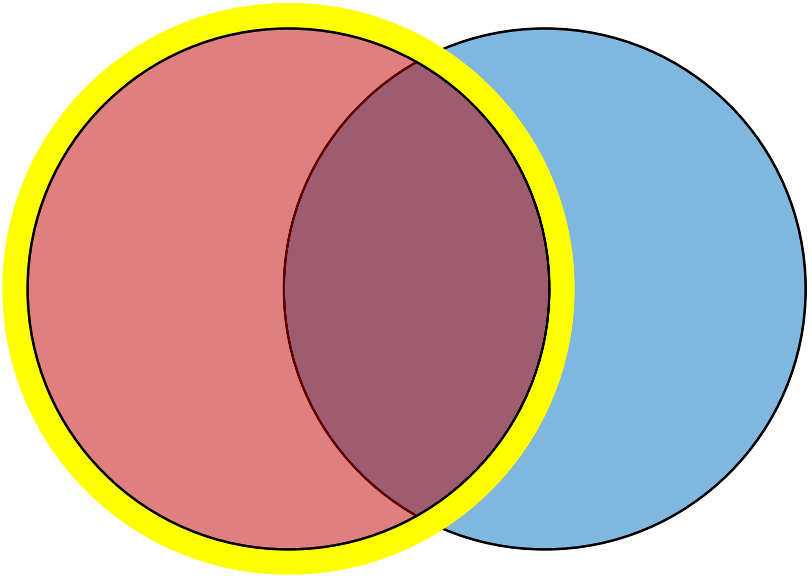
$$p(y|x) = \frac{p(x|y)p(y)}{p(x)}$$

  - (class) prior: $p(y)$
  - total

$$p(x) = p(x|y=0)p(y=0) + p(x|y=1)p(y=1)$$

Probability basics

•



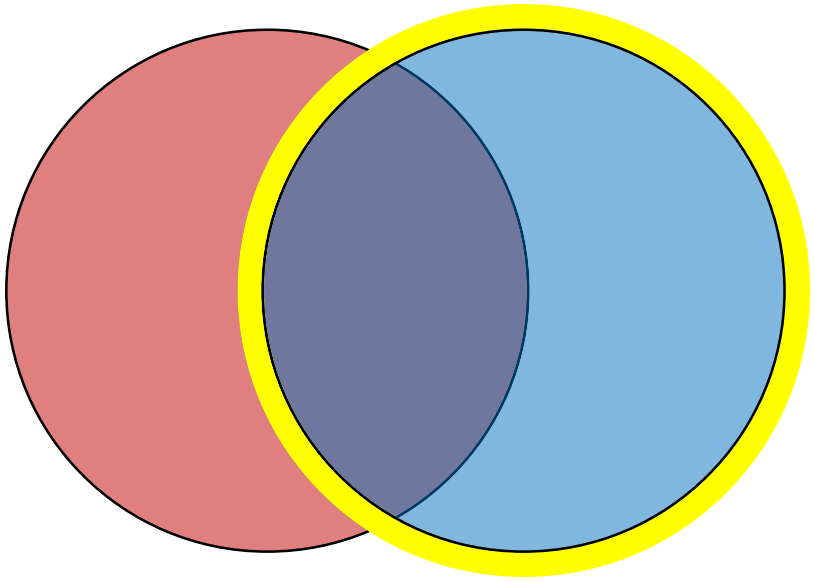
$P(X)$

- Single event probability:

$$P(X)$$

Probability basics (continued)

-

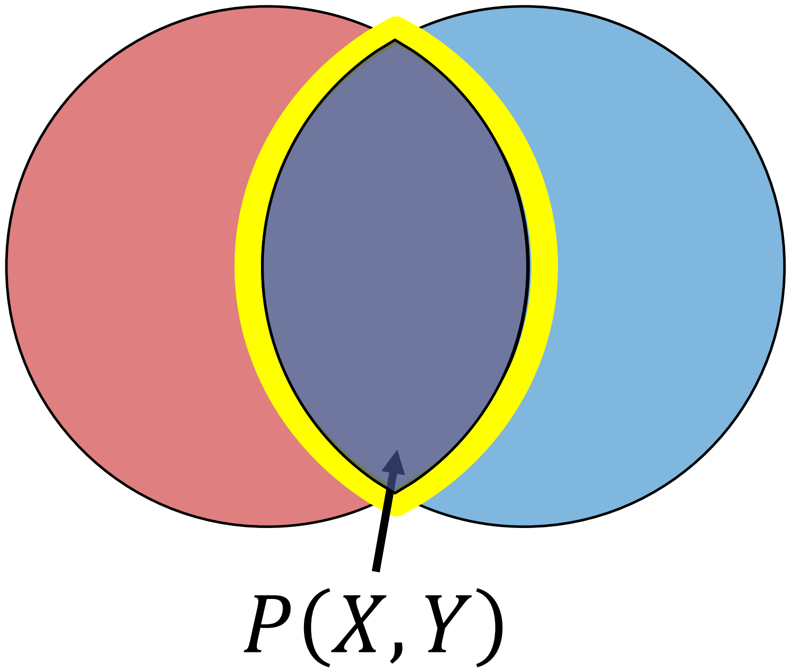


- Single event probability:

  $P(X), P(Y)$

Probability basics (continued)

•



- Single event probability:

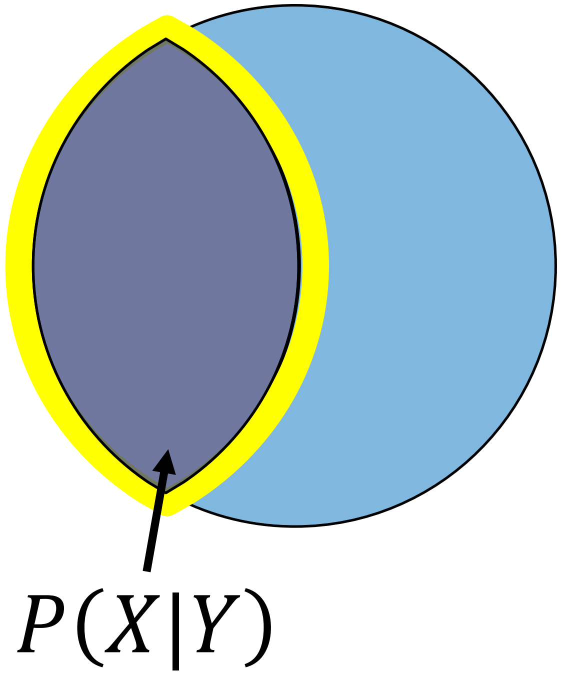
$$P(X), P(Y)$$

- Joint event probability:

$$P(X,Y)$$

Probability basics (continued)

•



- Single event probability:

$$P(X), P(Y)$$

- Joint event probability:

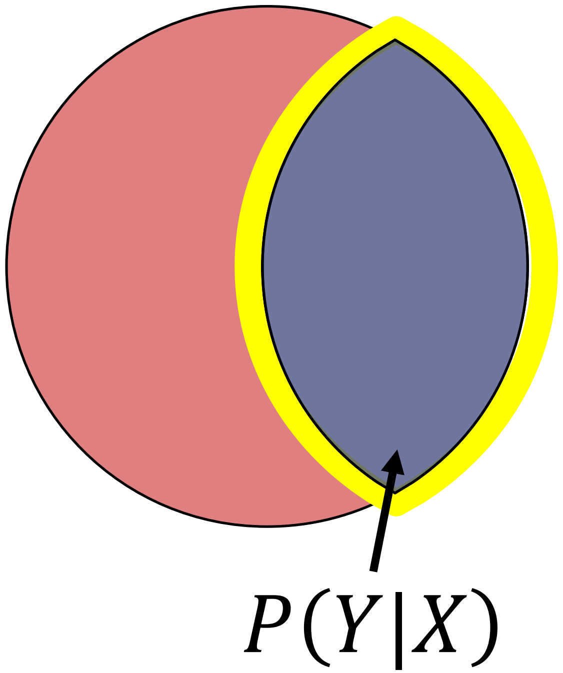
$$P(X, Y)$$

- Conditional probability:

$$P(X|Y)$$

Probability basics (continued)

-



- Single event probability:

$$P(X), P(Y)$$

- Joint event probability:

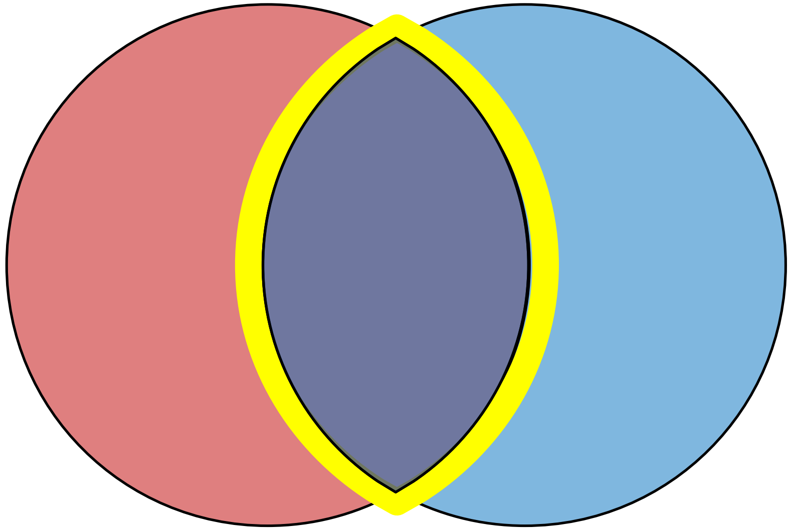
$$P(X, Y)$$

- Conditional probability:

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

Probability basics (continued)

-



- Single event probability:

$$P(X), P(Y)$$

- Joint event probability:

$$P(X, Y)$$

- Conditional probability:

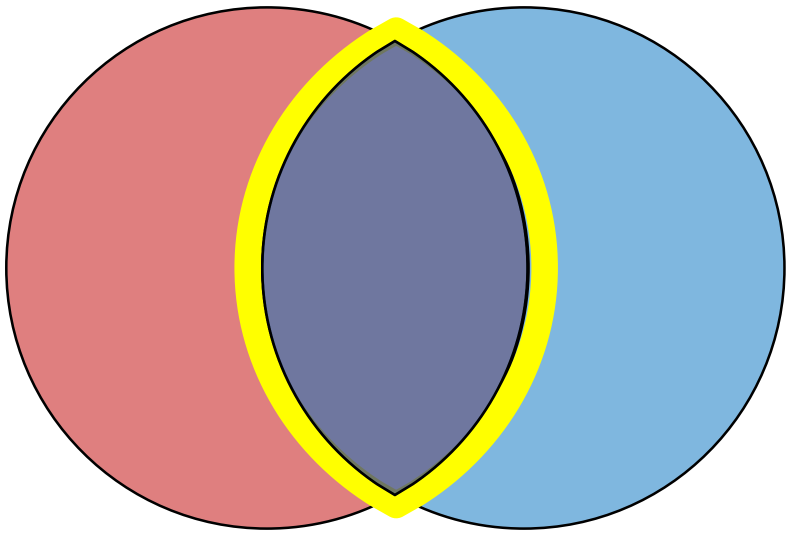
$$P(X|Y), P(Y|X)$$

- Joint and conditional relationship:

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

Bayes theorem derivation

•

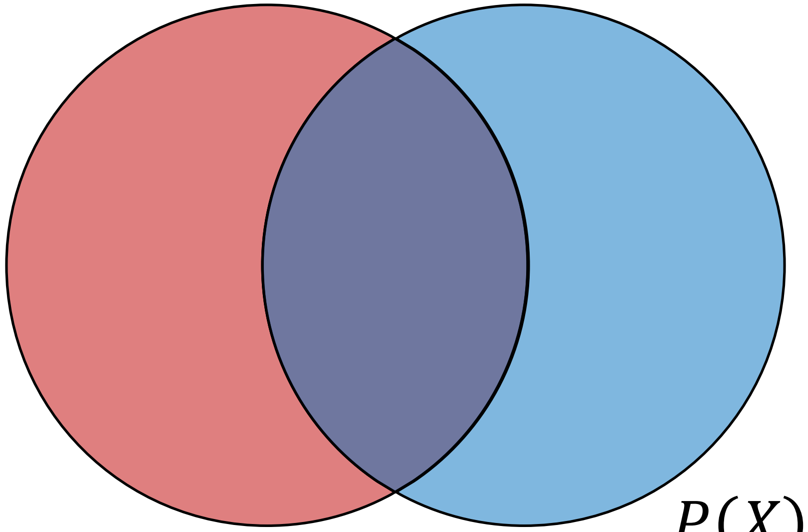


- By conditional and joint relationship:

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

Bayes theorem derivation (continued)

•



- Use conditional and joint relationship:

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

- To invert conditional probability:

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

$$P(X) = \sum_{Z} P(X,Z) = \sum_{Z} P(X|Z) * P(Z)$$

Your first generative learning algorithm

- Assume $p(x|y)$ is distributed according to a multivariate normal distribution
  - Multivariate normal distribution
    - Notebook

Gaussian discriminant analysis model

- Binary classification
  - Class prior

$$y \sim \text{Bernoulli}(\pi)$$

$$p(y) = \pi^y (1-\pi)^{(1-y)}$$

  - Likelihood

$$p(x|y=0) = \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}}} e^{\left(-\frac{1}{2}(x-\mu_0)^\intercal \Sigma^{-1} (x-\mu_0)\right)}$$

$$p(x|y=1) = \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}}} e^{\left(-\frac{1}{2}(x-\mu_1)^\intercal \Sigma^{-1} (x-\mu_1)\right)}$$

Gaussian discriminant analysis model (continued)

- Log-likelihood:

$$\ell(\pi, \mu_0, \mu_1, \Sigma) = \log \prod_{i=1}^{n_{\mathrm{train}}} p(x^{(i)}, y^{(i)}; \pi, \mu_0, \mu_1, \Sigma)$$

$$= \log \prod_{i=1}^{n_{\mathrm{train}}} p(x^{(i)} | y^{(i)}; \mu_0, \mu_1, \Sigma) p(y^{(i)}; \pi)$$

Gaussian discriminant analysis model (continued)

- Parameter estimate:
  - Parameter for prior

$$\pi = \frac{\sum_{i=1}^{n_{\text{train}}} \mathbb{1}[y^{(1)} = 1]}{n_{\text{train}}}$$

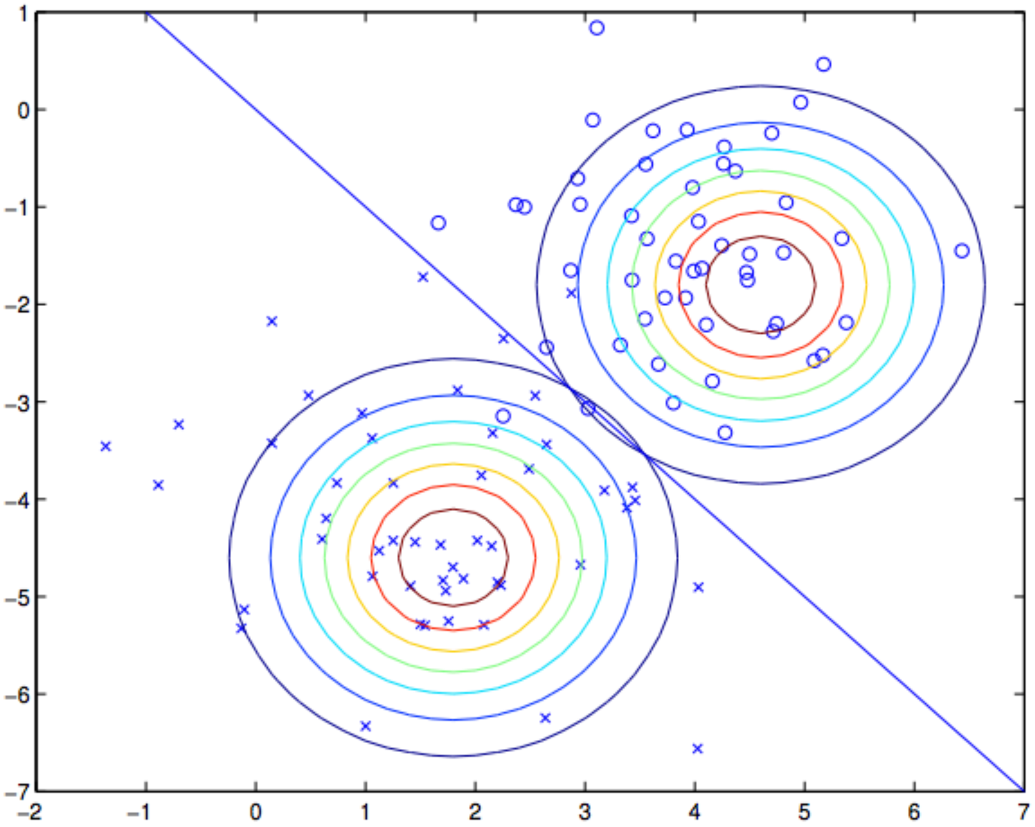
  - Parameters for posterior

$$\mu_0 = \frac{\sum_{i=1}^{n_{\text{train}}} \mathbb{1}[y^{(1)} = 0]x^{(i)}}{\sum_{i=1}^{n_{\text{train}}} \mathbb{1}[y^{(1)} = 0]} \qquad \mu_1 = \frac{\sum_{i=1}^{n_{\text{train}}} \mathbb{1}[y^{(1)} = 1]x^{(i)}}{\sum_{i=1}^{n_{\text{train}}} \mathbb{1}[y^{(1)} = 1]}$$

$$\Sigma = \frac{1}{n_{\text{train}}} \sum_{i=1}^{n_{\text{train}}} (x^{(i)} - \mu_{y^{(i)}})(x^{(i)} - \mu_{y^{(i)}})^{\mathsf{T}}$$

GDA model

- Decision boundary – linear discriminant analysis



GDA model – quadratic discriminant analysis

- LDA and QDA