Classification — Generative & Discriminative

Saeed Saremi

Assigned reading: 5.1, 5.2.1, 5.2.2, 5.2.4, 5.3

September 17, 2024

classification

- In contrast to the regression problem, the output is not a real number, but a label:

$$\mathcal{X} \to \mathcal{Y} = \{0, \dots, K-1\}$$

- The labels can be binary, e.g.

$$\mathbb{R}_+ = \{\text{cholesterol levels}\} \to \{0, 1\}$$

$$\{\text{proteins}\} \times \{\text{proteins}\} \to \{0, 1\}$$

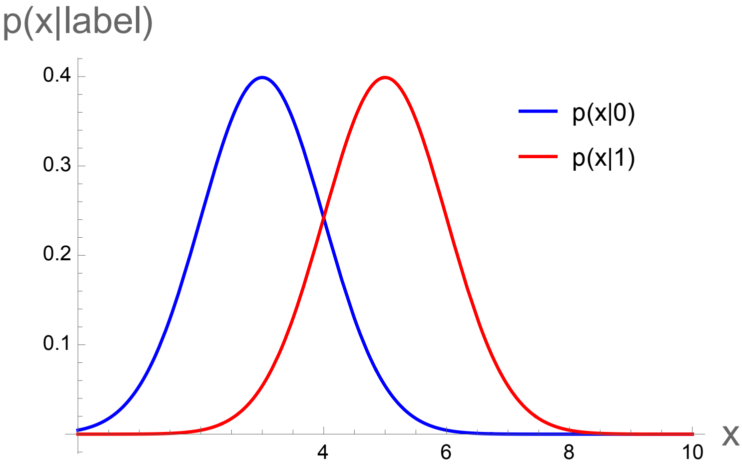
- The labels may not be binary, e.g. the MNIST handwritten digit classification:

$$\{\text{images}\} \to \{0, 1, \dots, 9\}$$

- Given a dataset $\{(x_i, y_i)\}_{i=1}^n$, we are interested in learning the mapping:

$$f_\theta : \mathcal{X} \to \{0, \dots, K-1\}$$

class-conditional probabilities



Here the class-conditional probabilities is assumed to be known (they can be estimated from data).

Bayes Rule

Bayes rule is used to go from class-conditional probabilities to the posterior probabilites

$$p(0|x) = \frac{p(x|0)p(0)}{p(x)},$$

$$p(1|x) = \frac{p(x|1)p(1)}{p(x)},$$

where

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

The prior probabilities can also be estimated from data, e.g. $p(0) = 2/3$

posterior probabilities