## **Optimal Classifier**

- 1. What's the optimal classifier when we can access infinite amount of data?
- 2. The Bayes rule based on P(Y=y | X=x)

## **Optimal Classifier**

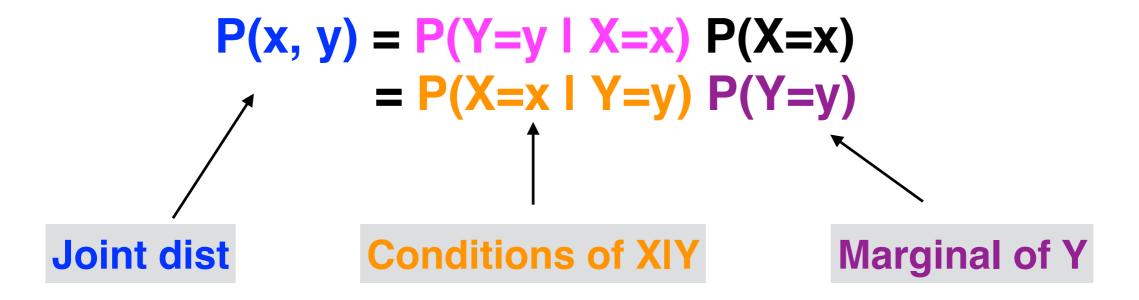
- 1. What's the optimal classifier when we can access infinite amount of data?
- 2. The Bayes rule based on P(Y=y | X=x)

$$P(x, y) = P(Y=y \mid X=x) P(X=x)$$

$$= P(X=x \mid Y=y) P(Y=y)$$

## **Optimal Classifier**

- 1. What's the optimal classifier when we can access infinite amount of data?
- 2. The Bayes rule based on P(Y=y | X=x)



### **Optimal Classifier**

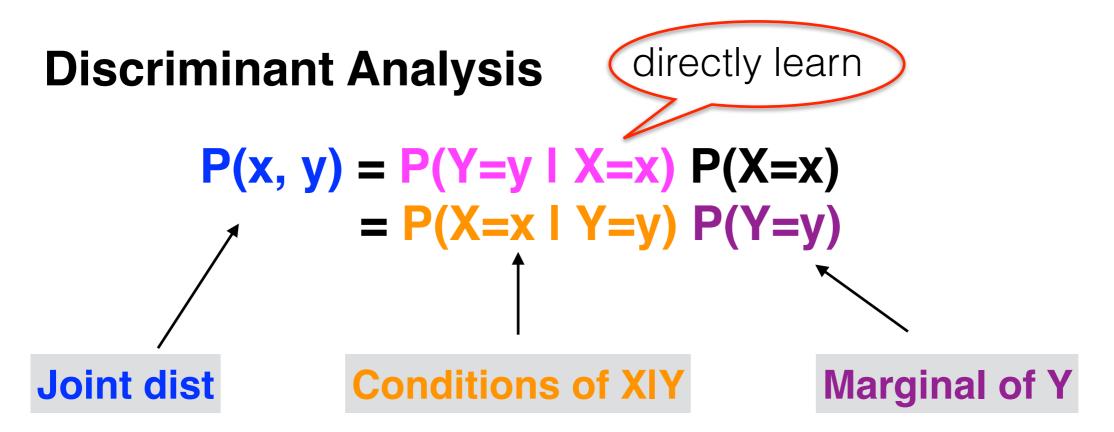
- 1. What's the optimal classifier when we can access infinite amount of data?
- 2. The Bayes rule based on P(Y=y | X=x)

#### **Discriminant Analysis**

Dist of p-dim X given Y=k: QDA, LDA (FDA), NB

## **Optimal Classifier**

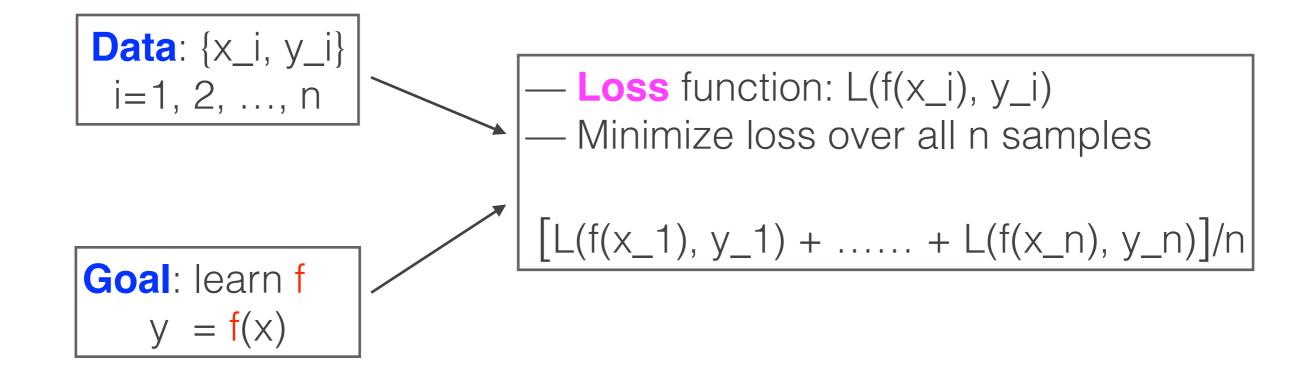
- 1. What's the optimal classifier when we can access infinite amount of data?
- 2. The Bayes rule based on P(Y=y | X=x)



Dist of p-dim X given Y=k: QDA, LDA (FDA), NB

#### How to Learn a Classifier

Focus on binary classification: y\_i = 0 or 1



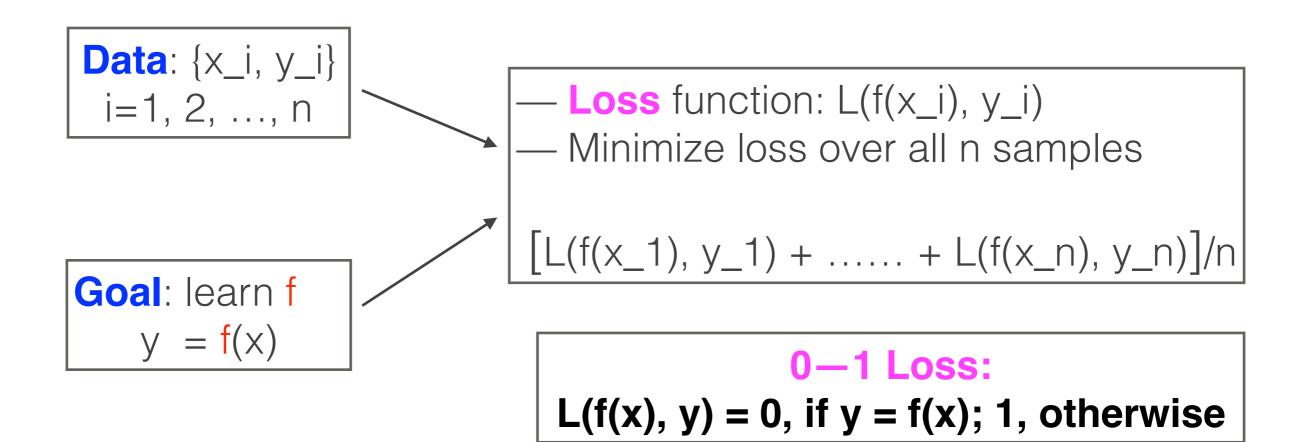
What's f?

input for f: p-dim feature input

output for f: 0/1, prob in [0, 1], or a real number

#### How to Learn a Classifier

Focus on binary classification: y\_i = 0 or 1



What's f?

input for f: p-dim feature input

output for f: **0/1**, prob in [0, 1], or a real number

# The Optimal Classifier

Consider an **ideal** situation: we have **infinite** samples (or equivalently, we know how data (x,y) are generated) and are allowed to use **any classifier**.

$$\frac{1}{n} \sum_{i=1}^{n} L(f(x_i), y_i) \to \mathbb{E}_{X,Y} L(f(X), Y),$$

$$\int_{\mathcal{X}} \int_{\mathcal{Y}} L(y, f(x)) p(x, y) dy dx$$

$$= \int_{\mathcal{X}} \int_{\mathcal{Y}} L(y, f(x)) p(y|x) p(x) dy dx$$

$$= \int_{\mathcal{X}} \left[ \int_{\mathcal{Y}} L(y, f(x)) p(y|x) dy \right] p(x) dx$$

# **The Optimal Classifier**

Consider an **ideal** situation: we have **infinite** samples (or equivalently, we know how data (x,y) are generated) and are allowed to use **any classifier**.

$$\frac{1}{n} \sum_{i=1}^{n} L(f(x_i), y_i) \to \mathbb{E}_{X,Y} L(f(X), Y),$$

For each x, find the optimal value of f(x) that minimizes the inside integral

$$\int_{\mathcal{X}} \int_{\mathcal{Y}} L(y, f(x)) p(x, y) dy dx$$

$$= \int_{\mathcal{X}} \int_{\mathcal{Y}} L(y, f(x)) p(y|x) p(x) dy dx$$

$$= \int_{\mathcal{X}} \left[ \int_{\mathcal{Y}} L(y, f(x)) p(y|x) dy \right] p(x) dx$$

# **The Optimal Classifier**

Consider an **ideal** situation: we have **infinite** samples (or equivalently, we know how data (x,y) are generated) and are allowed to use **any classifier**.

$$\frac{1}{n} \sum_{i=1}^{n} L(f(x_i), y_i) \to \mathbb{E}_{X,Y} L(f(X), Y),$$

Note that the inside integral is actually a summation since **Y** is discreet

$$\int_{\mathcal{X}} \int_{\mathcal{Y}} L(y, f(x)) p(x, y) dy dx$$

$$= \int_{\mathcal{X}} \int_{\mathcal{Y}} L(y, f(x)) p(y|x) p(x) dy dx$$

$$= \int_{\mathcal{X}} \left[ \int_{\mathcal{Y}} L(y, f(x)) p(y|x) dy \right] p(x) dx$$