Classification: Discriminant Analysis

Optimal Classifier

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

Classification: Discriminant Analysis

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)**

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

Classification: Discriminant Analysis

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)**

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

Joint dist → $P(x, y)$

Conditions of X|Y → $P(X=x \mid Y=y)$

Marginal of Y → $P(Y=y)$

Classification: Discriminant Analysis

Optimal Classifier

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

Discriminant Analysis

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

Joint dist — $P(x, y)$

Conditions of X|Y — $P(X=x \mid Y=y)$

Marginal of Y — $P(Y=y)$

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

Classification: Discriminant Analysis

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

directly learn

$$
\begin{aligned}
P(x, y) &= P(Y=y \mid X=x)\, P(X=x) \\
&= P(X=x \mid Y=y)\, P(Y=y)
\end{aligned}
$$

Joint dist **Conditions of X|Y** **Marginal of Y**

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

Data: $\{x_i, y_i\}$
i=1, 2, …, n

Goal: learn f
y = f(x)

— **Loss** function: $L(f(x_i), y_i)$
— Minimize loss over all n samples

$[L(f(x_1), y_1) + \ldots\ldots + L(f(x_n), y_n)]/n$

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

Data: $\{x_i, y_i\}$
$i=1, 2, \ldots, n$

Goal: learn f
$y = f(x)$

— **Loss** function: $L(f(x_i), y_i)$
— Minimize loss over all n samples

$$[L(f(x_1), y_1) + \ldots\ldots + L(f(x_n), y_n)]/n$$

0—1 Loss:
L(f(x), y) = 0, if y = f(x); 1, otherwise

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**.

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

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

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**.

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

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

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

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

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

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**.

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

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

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

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

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