Lecture 11/12: Generative Models for Classification

Tuesday, November 12, 2019 2:49 PM

# Course so far

ERM

- Learning theory
- Decision Trees
- Perceptron
- SVM
- Neural Nets

Discriminative Conditional

- Logistic Regression
- Ridge Regression

---

## Bayes Decision Rule

$P(Y=1 \mid X=x) = 0.7$ → predict $y=1$, then Prob of error $0.3$

$P(Y=-1 \mid X=x) = 0.3$ → predict $y=-1$, then " " " $0.7$

$$\rightarrow Err_P(h_{bayes}) = \sum_{x \in X} P(X=x) \cdot \min_{y \in Y} \left(1 - P(Y=y \mid X=x)\right)$$

↳ Bayes error rate → best possible error rate

---

## Generative Models

$$h(x) = \underset{y \in Y}{\arg\max} \{P(Y=y \mid X=x)\}$$

$$= \underset{y \in Y}{\arg\max} \{P(Y=y \mid X=x) \cdot P(X=x)\}$$

$$= \underset{y \in Y}{\arg\max} \{P(Y=y, X=x)\}$$

$$= \underset{y \in Y}{\arg\max} \{\underline{P(X=x \mid Y=y) \cdot P(Y=y)}\}$$

Idea:

- select distributions to consider for $P(X,Y)$
- find the distribution that best fits the training data
- use Bayes decision rule to derive a classifier

Intuition:

$P(Y=y)$: How likely is each class a priori?

$P(X=x \mid Y=1)$: How likely do I see $x$ in class $1$?

$P(X=x \mid Y=-1)$: How likely do I see $x$ in class $-1$?

Two model:

$$P(Y=1 \mid X=x) = \frac{P(X=x \mid Y=1) \cdot P(Y=1)}{P(X=x)}$$

$$P(Y=-1 \mid X=x) = \frac{P(X=x \mid Y=-1) \cdot P(Y=-1)}{P(X=x)}$$

---

Multivariate Naive Bayes

## Multivariate Naive Bayes

Assumption:

$$P(X=\vec{x} \mid Y=1) = P(X_1=x_1, X_2=x_2, \dots, X_N=x_N \mid Y=1) = P(x_1=x_1 \mid Y=1) \cdot \ldots \cdot P(x_N=x_N \mid Y=1)$$

$$P(X=\vec{x} \mid Y=-1) = \ldots$$

Estimation

$$\hat{P}(Y=1) = \frac{3}{4} \qquad \hat{P}(Y=-1) = \frac{1}{4}$$

$$\hat{P}(x_{fever}=high \mid Y=1) = \frac{2}{3} \qquad \hat{P}(x_{fever}=low \mid Y=1) = \frac{1}{3} \qquad \hat{P}(x_{fever}=no \mid Y=1) = 0$$

$$\hat{P}(x_{fever}=high \mid Y=-1) = 0 \qquad \hat{P}(x_{fever}=low \mid Y=-1) = 1 \qquad \hat{P}(x_{fever}=no \mid Y=-1) = 0$$

$\vdots$

Classification

$$P(Y=1, x=(high, no, yes)) = P(Y=1) \cdot P(x_{fever}=high \mid Y=1) \cdot P(x_{cough}=no \mid Y=1) \cdot P(x_{poke}=yes \mid Y=1)$$

$$= \frac{3}{4} \cdot \frac{2}{3} \cdot \frac{1}{3} \cdot \frac{2}{3}$$

$$= \cancel{\frac{1}{11}} \; \frac{1}{9}$$

$$P(Y=-1, x=(high, no, yes)) = P(Y=-1) \cdot P(x_{fever}=high \mid Y=-1) \cdot P(x_{cough}=no \mid Y=-1) \cdot P(x_{poke}=yes \mid Y=-1)$$

$$= \frac{1}{4} \cdot 0$$

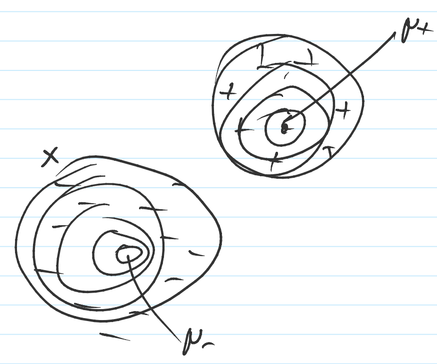
$$= 0$$

## Linear Discriminant Analysis (LDA)

$$h(x) = \arg\max_y \{ P(X=x \mid Y=y) \cdot P(Y=y) \}$$

$$P(X=x \mid Y=1) = N(\vec{\mu}_+, 1) = \frac{1}{Z} e^{-\frac{1}{2}(x-\mu_+)^2}$$

$$P(X=x \mid Y=-1) = N(\vec{\mu}_-, 1) = \frac{1}{Z} e^{-\frac{1}{2}(x-\mu_-)^2}$$




$$\to \arg\max_y \{ N(\mu_y, 1) \cdot P(Y=y) \}$$

$$= \arg\max_y \{ e^{-\frac{1}{2}(\mu_y - x)^2} \cdot P(Y=y) \}$$

$$= \arg\max_y \{ -\frac{1}{2}(\mu_y - x)^2 + \log P(Y=y) \}$$