Universidad de Buenos Aires
Facultad de Ingenieria
Año 2018 - 2er Cuatrimestre

Aprendizaje Estadístico, Teoría y aplicación

Resúmen de la materia y devolutiva

Sbruzzi, José Ignacio - Ingeniería Informática #97452
jose.sbru@gmail.com

Índice

1. Clase 1 (31/8)

La primera parte de esta clase fue un repaso de diversos temas que serán útiles durante la cursada:

- Teorema de pitágoras
- espacios euclídeos
- Ortogonalidad
- Relación entre el producto interno y el coseno
- Desigualdad Cauchy-Schwartz
- Norma inducida
- Proyección Ortogonal
- Definición de esperanza
- El espacio algebraico de variables aleatorias
- Desigualdad de Markov
- Desigualdad de Chebyshev
- Desigualdad de Chernoff
- Desigualdad de Jensen
- Función convexa
- Esperanza condicional

La segunda parte de la clase se habló del problema de la comunicación digital para ilustrar la lógica por detrás de la construcción de un clasificador bayesiano. Siendo $\delta(r)$ una función que predice el dígito (0 o 1) emitido a partir del recibido $r \in \{0, 1\}$. $P(S = s|R = r)$ es la probabilidad de que se haya emitido el dígito s dado que se recibió el dígito r.

$$\delta(r) = \mathbb{1}\{\mathbb{P}(S = 1|R = r) > \mathbb{P}(S = 0|R = r)\}$$

Así, este clasificador toma la mejor decisión posible para la información que se tiene disponible (r), con lo cual es un clasificador bayesiano.

2. Clase 2 (7/9)

2.1. Definiciones iniciales

2.1.1. Clasificador

$$g : \mathbb{R}^d \to \{1, 2, ..., M\}$$

$g(x)$ representa una conjetura respecto de la naturaleza de la distribución de las x. El clasificador se equivoca cuando $g(x) \neq y$.

2.1.2. Calidad del clasificador

Sea $(X, Y) \in \mathbb{R}^d \times \{1, 2, ..., M\}$ un par donde X es una variable aleatoria que representa las propiedades observables y Y la característica a predecir. Así, se define la pérdida de un clasificador como $L(g) = \mathbb{P}(g(X) \neq Y)$.

2.1.3. Clasificador bayesiano

Es el mejor clasificador, definido por

$$argmin_{g:\mathbb{R}^d \to \{1,...,M\}}\{\mathbb{P}(g(X) \neq Y)\} = g^*$$

$$L^* = L(g^*)$$

No se da siempre que $L^* = 0$ porque Y podría no ser una función de X.

2.1.4. Dataset de entrenamiento

Se denota como $(X_i, Y_i), i = 1, 2, ..., n$; donde las parejas (X_i, Y_i) son observaciones independientes e identicamente distribuidas, al igual que (X, Y).

$$D_n = \{(X_i, Y_i), i = 1, 2, ..., n\}$$

Así, en realidad, cuando aplicamos algoritmos de machine learning tenemos una g denotada como:

$$g(X, (X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n))$$

Donde X es una nueva observación.

Es decir,

$$g_n : \mathbb{R}^d \times (\mathbb{R}^d \times \{1, ..., M\})^n \to \{1, ..., M\}$$

Así, tenemos

$$L_n = L(g_n) = \mathbb{P}(g(X, (X_1, Y_1), ..., (X_n, Y_n)) \neq Y | (X_1, Y_1), ..., (X_n, Y_n))$$

Con lo cual L_n es una variable aleatoria dependiente de las observaciones.

2.2. Clasificador bayesiano para M=2

Sean (con $A \subset \mathbb{R}^d$, $x \in \mathbb{R}^d$, $y \in \{0, 1\}$):

$$\mu(A) = \mathbb{P}(x \in A)$$
$$\eta(x) = \mathbb{P}(Y = 1|X = x) = \mathbb{E}[y|X = x]$$

Así,

$$\eta(x) = \int_C \mathbb{P}(Y = 0|X = x)\mu(dx) + \int_C \mathbb{P}(Y = 1|X = x)\mu(dx)$$

Siendo $C = \mathbb{R}^d \times \{0, 1\}$.

Bajo estas condiciones,

$$g^*(x) = \mathbb{1}\{\eta(x) > \frac{1}{2}\}$$

3. Clase 3 (14/9)