Clasificación Supervisada

Métodos Estadísticos

Jesús Ariel Carrasco Ochoa
Instituto Nacional de Astrofísica, Óptica y Electrónica

Probabilidad

Probabilidad

$$P(X) = \frac{\#\,\text{de eventos a favor}}{\#\,\text{total de eventos posibles}}$$

Si ***X*** y ***Y*** son eventos excluyentes

$$P(X \vee Y) = P(X) + P(Y)$$

Probabilidad

Si ***X*** y ***Y*** son eventos independientes

$$P(X \wedge Y) = P(X)P(Y)$$

Si ***X*** y ***Y*** son eventos no independientes

$$P(X \wedge Y) = P(X)P(Y/X)$$

$$P(X \wedge Y) = P(Y)P(X/Y)$$

Probabilidad

Como tenemos 2 formas de calcular $P(X \wedge Y)$

$$P(Y)P(X/Y) = P(X)P(Y/X)$$

Despejando

$$P(X/Y) = \frac{P(X)P(Y/X)}{P(Y)}$$

Clasificación Bayesiana

- Sea T un conjunto de entrenamiento con n objetos $O_1,\dots,O_n$
- Los objetos de T están agrupados en 2 clases C_1 y C_2 y se conoce a la clase a la que pertenece cada objeto
- Sea n_1 el número de objetos de T en C_1 y n_2 el número de objetos de T en C_2
- Cada objeto O_j descrito por una sola variable x

Clasificación Bayesiana

Dado un nuevo objeto ***O***

- En el enfoque estadístico se parte de la idea de colocar un objeto nuevo en la clase que sea más probable

- La forma más simple de hacerlo es colocarlo en la clase mas probable sin tomar en cuanta la descripción del objeto
  - Debemos conocer las probabilidades de cada clase (si no se conocen usualmente se aproximan como $P(C_i)=n_i/n$

  - Todos los objetos nuevos se pondrán en la misma clase

Clasificación Bayesiana

Si el valor de ***O*** en ***x*** lo denotamos como ***x(O)***

- Para tomar en cuenta la descripción del objeto a clasificar debemos calcular las probabilidades condicionales

$$P(O \in C_i / x(O) = X) \text{ que denotaremos como } P(C_i / X)$$

- Y colocar al objeto ***O*** en la clase con mayor probabilidad

Clasificación Bayesiana

Cómo calcular $P(C_i / X)$?

$$P(C_i / X) = \frac{P(C_i)P(X / C_i)}{P(X)}$$

Donde $P(C_i)$ la aproximamos como

$$P(C_i) = \frac{n_i}{n}$$

Clasificación Bayesiana

Como queremos sólo saber cuál $P(C_i / X)$ es mayor podemos eliminar $P(X)$ y solamente calcular

$$P(C_i)P(X/C_i)$$

Si suponemos que $\boldsymbol{x} \sim N(\mu_i, \sigma_i^2)$ en C_i

$$P(X/C_i) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{1}{2}\frac{(X-\mu_i)^2}{\sigma_i^2}}$$

Clasificación Bayesiana

A ***O*** se le asigna la clase C_O si

$$C_O = \max_{C_i} \left\{ P(C_i) P(X / C_i) \right\}$$

Para $\boldsymbol{x} \sim N(\mu_i, \sigma_i^2)$ en C_i queda

$$C_O = \max_{C_i} \left\{ \frac{n_i}{n} \frac{1}{\sqrt{2\pi\sigma_i}} e^{-\frac{1}{2}\frac{(X-\mu_i)^2}{\sigma_i^2}} \right\}$$

Clasificación Bayesiana

Tomado logaritmo natural queda

$$C_O = \max_{C_i}\left\{ \ln\left(\frac{n_i}{n}\right) - \ln\left(\sqrt{2\pi}\sigma_i\right) - \frac{(X-\mu_i)^2}{2\sigma_i^2} \right\}$$

Clasificación Bayesiana

De donde salen $\boldsymbol{\mu_i}$ ***y*** $\boldsymbol{\sigma_i^2}$?

Se aproximan con la media y la desviación estándar de la muestra

$$\mu_i = \overline{X}_i = \frac{1}{n_i} \sum_{O_j \in C_i} x(O_j)$$

$$\sigma_i^2 = S_i^2 = \frac{1}{n_i - 1} \sum_{O_j \in C_i} \left(x(O_j) - \overline{X}_i \right)$$

Clasificación Bayesiana

- Sea T un conjunto de entrenamiento con $\boldsymbol{n}$ objetos $\boldsymbol{O_1,\ldots,O_n}$
- Los objetos de $\boldsymbol{T}$ están agrupados en r clases $\boldsymbol{C_1,\ldots,C_r}$ y se conoce a la clase a la que pertenece cada objeto
- Sea $\boldsymbol{n_i}$ el número de objetos de $\boldsymbol{T}$ en $\boldsymbol{C_i}$
- Cada objeto $\boldsymbol{O_j}$ descrito por $\boldsymbol{m}$ variables $\boldsymbol{x_1,\ldots,x_m}$

Clasificación Bayesiana

En este caso queremos calcular las probabilidades condicionales

$$P\left(O \in C_i \,/\, O = \left(x_1(O), \ldots, x_m(O)\right)\right)$$

Que denotaremos como

$$P(C_i \,/\, X)$$

Clasificación Bayesiana

Como queremos sólo saber cuál $P(C_i / X)$, solamente calculamos

$$P(C_i)P(X/C_i)$$

Si suponemos que $\boldsymbol{x} \sim N(\boldsymbol{\mu_i}, \boldsymbol{\Sigma_i})$ en C_i

$$P(X/C_i) = \frac{1}{\sqrt{(2\pi)^m |\Sigma_i|}} e^{-\frac{1}{2}(X-\mu_i)^T \Sigma_i^{-1} (X-\mu_i)}$$

Clasificación Bayesiana

A ***O*** se le asigna la clase C_O si

$$C_O = \max_{C_i} \left\{ P(C_i) P(X / C_i) \right\}$$

Para ***x***~*N(**μ**$_i$**, Σ**$_i$)* en C_i queda

$$C_O = \max_{C_i} \left\{ \frac{n_i}{n} \frac{1}{\sqrt{(2\pi)^m \left| \Sigma_i \right|}} e^{-\frac{1}{2}(X-\mu_i)^T \Sigma_i^{-1} (X-\mu_i)} \right\}$$

Clasificación Bayesiana

Tomado logaritmo natural queda

$$C_O = \max_{C_i}\left\{ \ln\left(\frac{n_i}{n}\right) - \ln\left(\sqrt{2\pi\left|\Sigma_i\right|}\right) - \tfrac{1}{2}(X-\mu_i)^T \Sigma_i^{-1}(X-\mu_i) \right\}$$

Clasificación Bayesiana

De dónde salen μ_i y Σ_i ?

Se aproximan con la media y la desviación estándar de la muestra

$$\mu_i = \overline{X_i} = \frac{1}{n_i} \sum_{O_j \in C_i} \left(x_1(O_j), \ldots, x_m(O_j) \right)$$

$$\Sigma_i = S_{ij} = \frac{1}{n_i - 1} \sum_{O_j \in C_i} \left(X_j - \overline{X_i} \right) \left(X_j - \overline{X_i} \right)^T$$

Clasificador Simple de Bayes (Naïve Bayes)

Para simplificar el trabajo se asume que las variables que describen a los objetos son independientes con lo que

$$P(X/C_i) = \prod_{j=1}^{m} P(X_j/C_i)$$

Clasificador Simple de Bayes (Naïve Bayes)

Si suponemos que $\boldsymbol{X_j} \sim N(\boldsymbol{\mu_{ij}}, \boldsymbol{\sigma_{ij}^2})$ en C_i

$$P(X/C_i) = \prod_{j=1}^{m} \left(\frac{1}{\sqrt{2\pi\sigma_{ij}}} e^{-\frac{1}{2}\frac{(X_j - \mu_{ij})^2}{\sigma_{ij}^2}} \right)$$

Clasificador Simple de Bayes (Naïve Bayes)

A ***O*** se le asigna la clase C_O si

$$C_O = \max_{C_i} \left\{ P(C_i) P(X / C_i) \right\}$$

Para $\boldsymbol{X_j} \sim N(\mu_{ij}, \sigma_{ij}^2)$ en C_i queda

$$C_O = \max_{C_i} \left\{ \frac{n_i}{n} \prod_{j=1}^{m} \left(\frac{1}{\sqrt{2\pi\sigma_{ij}}} e^{-\frac{1}{2}\frac{(X_j - \mu_{ij})^2}{\sigma_{ij}^2}} \right) \right\}$$

Clasificador Simple de Bayes (Naïve Bayes)

Tomado logaritmo natural queda

$$C_O = \max_{C_i}\left\{\ln\left(\frac{n_i}{n}\right) + \sum_{j=1}^{m}\left(-\ln\left(\sqrt{2\pi}\sigma_{ij}\right) - \frac{(X-\mu_{ij})^2}{2\sigma_{ij}^2}\right)\right\}$$

$$C_O = \max_{C_i}\left\{\ln\left(\frac{n_i}{n}\right) - \sum_{j=1}^{m}\ln\left(\sqrt{2\pi}\sigma_{ij}\right) - \sum_{j=1}^{m}\frac{(X-\mu_{ij})^2}{2\sigma_{ij}^2}\right\}$$

Estimación de la función de densidad

Se buscan los $\boldsymbol{k}$ vecinos más cercanos a $\boldsymbol{X}$ en $\boldsymbol{C_i}$

Sea $\boldsymbol{d_k(X)}$ la distancia entre $\boldsymbol{X}$ y el vecino más lejano de los $\boldsymbol{k}$ más cercanos

Sea $\boldsymbol{k_i}$ el número de objetos en $\boldsymbol{C_i}$ entre los más cercanos a $\boldsymbol{X}$

Estimación de la función de densidad

La función de densidad de ***X*** en ***C_i*** puede estimarse como:

$$\hat{P}(X/C_i) = \frac{k_i}{n_i V_k(X)}$$

donde

$$V_k(X) = \frac{\pi^{\frac{n}{2}}\left(d_k(X)\right)^n}{\Gamma\left(\frac{n}{2}+1\right)}$$

$$\Gamma\left(\frac{n}{2}+1\right) = \begin{cases} \left(\frac{n}{2}\right)! & \text{si n es par} \\ \left(\frac{n}{2}\right)\Gamma\left(\frac{n}{2}\right) & \text{si n es impar, con } \Gamma(0.5) = \sqrt{\pi} \end{cases}$$

Estimación de la función de densidad

Podemos analizar qué tan buena es la suposición de que los datos siguen una cierta distribución

Si suponemos que ***X*** sigue la distribución $f(X)$ en $\boldsymbol{C_i}$ entonces

$$P(X/C_i) = f(X)$$

Estimación de la función de densidad

Calcular

$$\chi_i^2 = n_i \sum_{j=1}^{n_i} \frac{\left(\hat{P}(X_j / C_i) - P(X_j / C_i)\right)^2}{P(X_j / C_i)}$$

- Con $\boldsymbol{n_i\text{-}s\text{-}1}$ grados de libertad
- $\boldsymbol{s}$ es el número de parámetros estimados
- Se debe cumplir que $\boldsymbol{n_i \geq s+1}$

Estimación de la función de densidad

Si suponemos $X \sim N(\mu_i, \Sigma_i)$ en C_i

n medias

$\frac{n(n+1)}{2}$ varianzas

$$s = n + \frac{n(n+1)}{2} = \frac{n(n+3)}{2}$$

Estimación de la función de densidad

Se define el porcentaje de riesgo

Se checa en la tabla de χ^2 si el valor calculado es mayor que el máximo esperado

SI → Se rechaza la hipótesis

NO → No se rechaza ("se acepta") la hipótesis

Clasificación Supervisada

Métodos Estadísticos

Jesús Ariel Carrasco Ochoa
Instituto Nacional de Astrofísica, Óptica y Electrónica