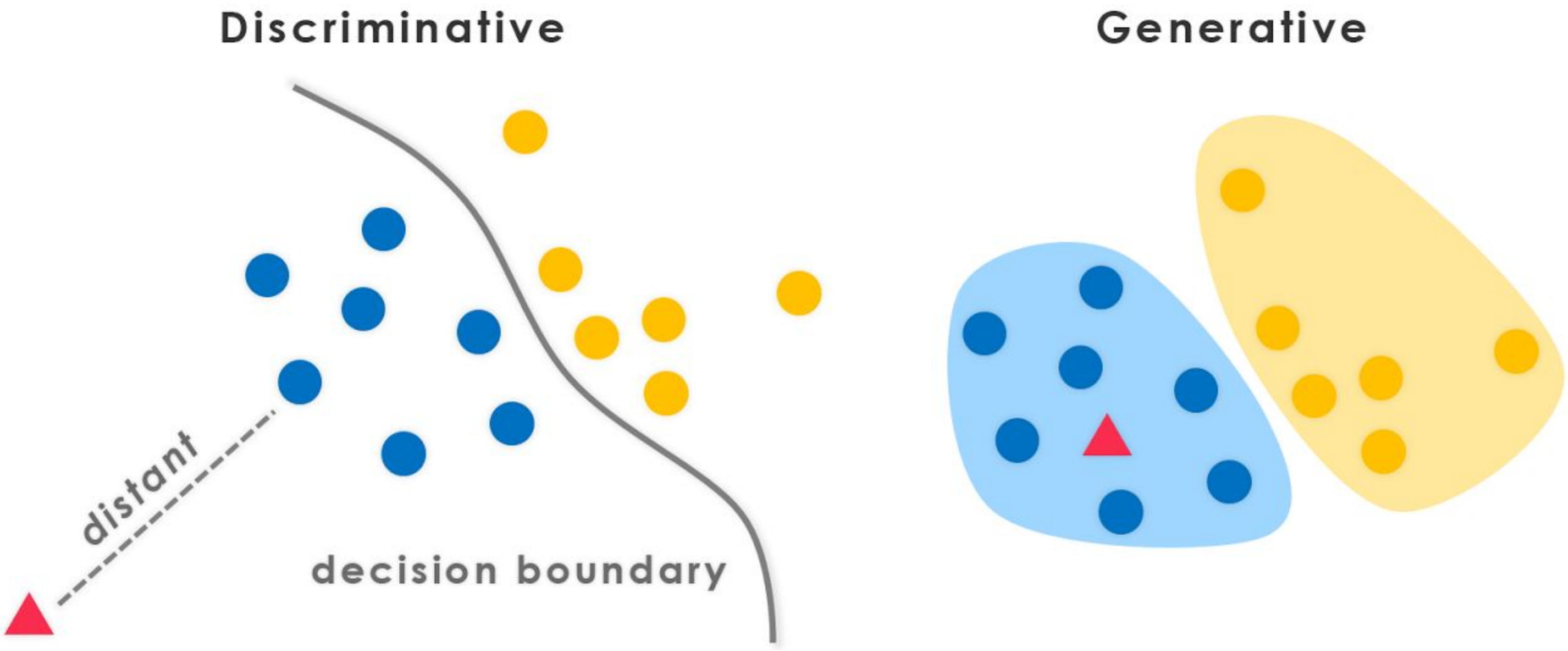
Naive Bayes _

Motivación

Generativos vs Discriminativos

- Existen variadas formas de implementar modelos de clasificación.
- Hasta el momento hemos trabajado con clasificadores discriminativos como la regresión logística.
- Estos aprenden las clases a inferir mediante una función objetivo de X a y.
- Ahora aprenderemos sobre clasificadores generativos, que aprenden las clases a inferir mediante la probabilidad conjunta de los atributos.

Generativos vs Discriminativos



Fuente: https://datawarrior.wordpress.com/2016/05/08/generative-discriminative-pairs/

Teorema de Bayes

Intuición

Se puede entender en base a la noción de probabilidad condicional.

Objetivo: Analizar la probabilidad de ocurrencia del evento condicionante B, dado que ocurra el evento condicionado.

- El teorema de Bayes también se conoce como probabilidad inversa.
- **Principal ventaja**: **Permite incorporar información adicional al análisis.**

Intuición

$$\Pr(\text{A posteriori}) = \frac{\Pr(\text{Verosimilitud}) \times \Pr(\text{A priori})}{\Pr(\text{Evidencia})}$$

Mantra Bayesiano: **La probabilidad a posteriori de un fenómeno es proporcional a la verosimilitud por la probabilidad a priori, ajustada por la evidencia.**

Intuición

$$\mathsf{Pr}(\text{A posteriori}) = \frac{\mathsf{Pr}(\text{Verosimilitud}) \times \mathsf{Pr}(\text{A priori})}{\mathsf{Pr}(\text{Evidencia})}$$

A posteriori: ¿Qué necesitamos saber?

Intuición

$$\mathrm{Pr}(\text{A posteriori}) = \frac{\mathrm{Pr}(\text{Verosimilitud}) \times \mathrm{Pr}(\text{A priori})}{\mathrm{Pr}(\text{Evidencia})}$$

Verosimilitud: La información que asumimos y a la cual tenemos conocimiento sobre su ocurrencia.

Intuición

$$\Pr(\text{A posteriori}) = \frac{\Pr(\text{Verosimilitud}) \times \Pr(\text{A priori})}{\Pr(\text{Evidencia})}$$

A priori: La probabilidad previa de ocurrencia de nuestro evento de interés. Puede afectar de manera substancial a la verosimilitud.

Intuición

$$\text{Pr}(\text{A posteriori}) = \frac{\text{Pr}(\text{Verosimilitud}) \times \text{Pr}(\text{A priori})}{\text{Pr}(\text{Evidencia})}$$

Evidencia: Evidencia existente sobre el fenómeno condicionante.

Bayes Ingenuo

Intuición

$$\hat{y}_{\text{map}} = \underset{y \in \mathbb{Y}}{\operatorname{argmax}}\ \hat{\Pr}(y|X) = \underset{y \in \mathbb{Y}}{\operatorname{argmax}}\ \hat{\Pr}(y) \prod_{1 \leq k \leq n_d} \Pr(X_k|y)$$

La fórmula de Bayes Ingenuo se puede reexpresar de la siguiente manera.
El objetivo es encontrar un **máximo a posteriori**: Dado que no tenemos suficiente información como para asumir las probabilidades como verdaderas, optamos por elegir la probabilidad de clase más alta entre todas.
El problema de optimización responde a argmax en el cálculo de todas las **probabilidades de clase.**

Intuición: A priori

$$\hat{y}_{\text{map}} = \underset{y \in \mathbb{Y}}{\operatorname{argmax}} \ \hat{\Pr}(y|X) = \underset{y \in \mathbb{Y}}{\operatorname{argmax}} \ \hat{\Pr}(y) \prod_{1 \leq k \leq n_d} \Pr(X_k|y)$$

Intuición: A priori

$$\hat{y}_{\mathrm{map}} = \underset{y \in \mathbb{Y}}{\operatorname{argmax}} \; \hat{\Pr}(y|X) = \underset{y \in \mathbb{Y}}{\operatorname{argmax}} \; \hat{\Pr}(y) \prod_{1 \leq k \leq n_d} \Pr(X_k|y)$$

$$\hat{\Pr}(y) = \frac{N_{y \in \mathbb{Y}}}{N}$$

Intuición: A priori

$$\hat{y}_{\text{map}} = \underset{y \in \mathbb{Y}}{\operatorname{argmax}} \ \hat{\Pr}(y|X) = \underset{y \in \mathbb{Y}}{\operatorname{argmax}} \ \hat{\Pr}(y) \prod_{1 \le k \le n_d} \Pr(X_k|y)$$

$$\hat{\Pr}(y) = \frac{N_{y \in \mathbb{Y}}}{N}$$

A priori no informativo: Probabilidad de ocurrencia uniforme para cada clase.

En caso de modificar el a priori, éste debe reflejar la cantidad de clases, y respetar los Axiomas de Kolmogorov.

Intuición: Verosimilitud de eventos

$$\hat{y}_{\text{map}} = \underset{y \in \mathbb{Y}}{\operatorname{argmax}} \hat{\Pr}(y|X) = \underset{y \in \mathbb{Y}}{\operatorname{argmax}} \hat{\Pr}(y) \prod_{1 \leq k \leq n_d} \Pr(X_k|y)$$

Intuición: Verosimilitud de eventos

$$\hat{y}_{\text{map}} = \operatorname*{argmax}_{y \in \mathbb{Y}} \hat{\Pr}(y|X) = \operatorname*{argmax}_{y \in \mathbb{Y}} \hat{\Pr}(y) \prod_{1 \leq k \leq n_d} \Pr(X_k|y)$$

$$\hat{\Pr}(X_k|y) = \frac{X_{yk}}{\sum X_{yk}}$$

Intuición: Verosimilitud de eventos

$$\hat{y}_{\text{map}} = \underset{y \in \mathbb{Y}}{\operatorname{argmax}}\ \hat{\Pr}(y|X) = \underset{y \in \mathbb{Y}}{\operatorname{argmax}}\ \hat{\Pr}(y) \prod_{1 \leq k \leq n_d} \Pr(X_k|y)$$

$$\hat{\Pr}(X_k|y) = \frac{X_{yk}}{\sum X_{yk}}$$

La verosimilitud de ocurrencia se obtiene a partir de la frecuencia relativa entre una clase y un atributo.

Independencia Condicional: el algoritmo es ingenuo dado que asumimos que la presencia/ausencia de una clase no depende de la presencia/ausencia de un atributo específico.

Entrenamiento

$$\hat{y}_{\text{map}} = \underset{y \in \mathbb{Y}}{\operatorname{argmax}} \left[\log \hat{\Pr}(y) + \sum_{1 \leq k \leq n_d} \log \hat{\Pr}(X_k | y) \right]$$

La fórmula descrita anteriormente presenta un inconveniente: **Estamos trabajando con la productoria.** Por lo general, preferimos convertir la productoria en la sumatoria del logaritmo, en un truco que se conoce como el **log-exp-sum** (Murphy, 2012).

{desafío} latam_

Academia de talentos digitales

www.desafiolatam.com