



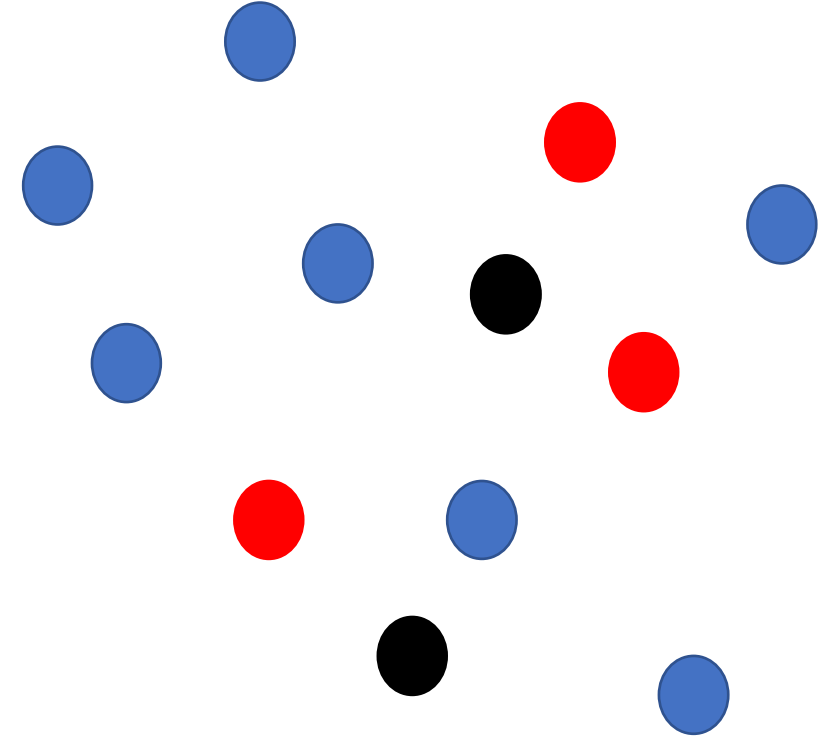
KNN – Naïve Bayes



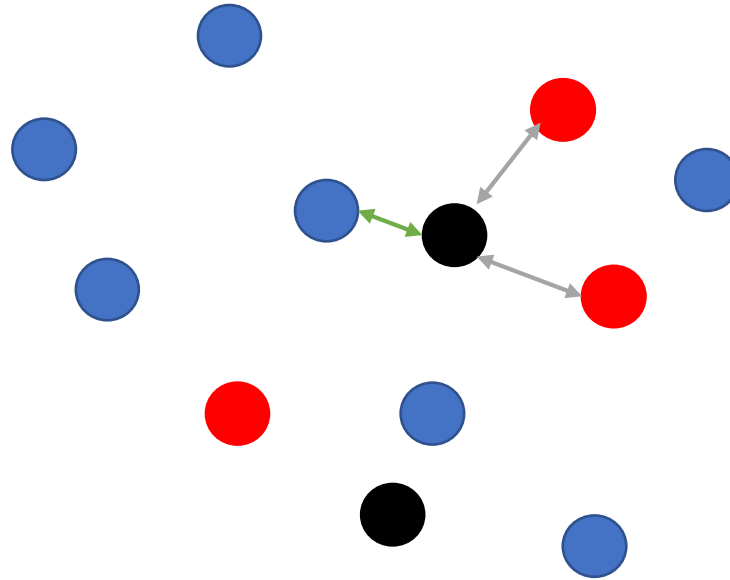
K – Nearest Neighbors

Idea Fundamental

- La clasificación de un caso está dada por la clasificación encontrada en sus casos más cercanos.
- La escogencia del número de casos (k) es un hiperparámetro a sintonizar.
- También puede usarse para estimar (promedio de los k más cercanos).



Ejemplo



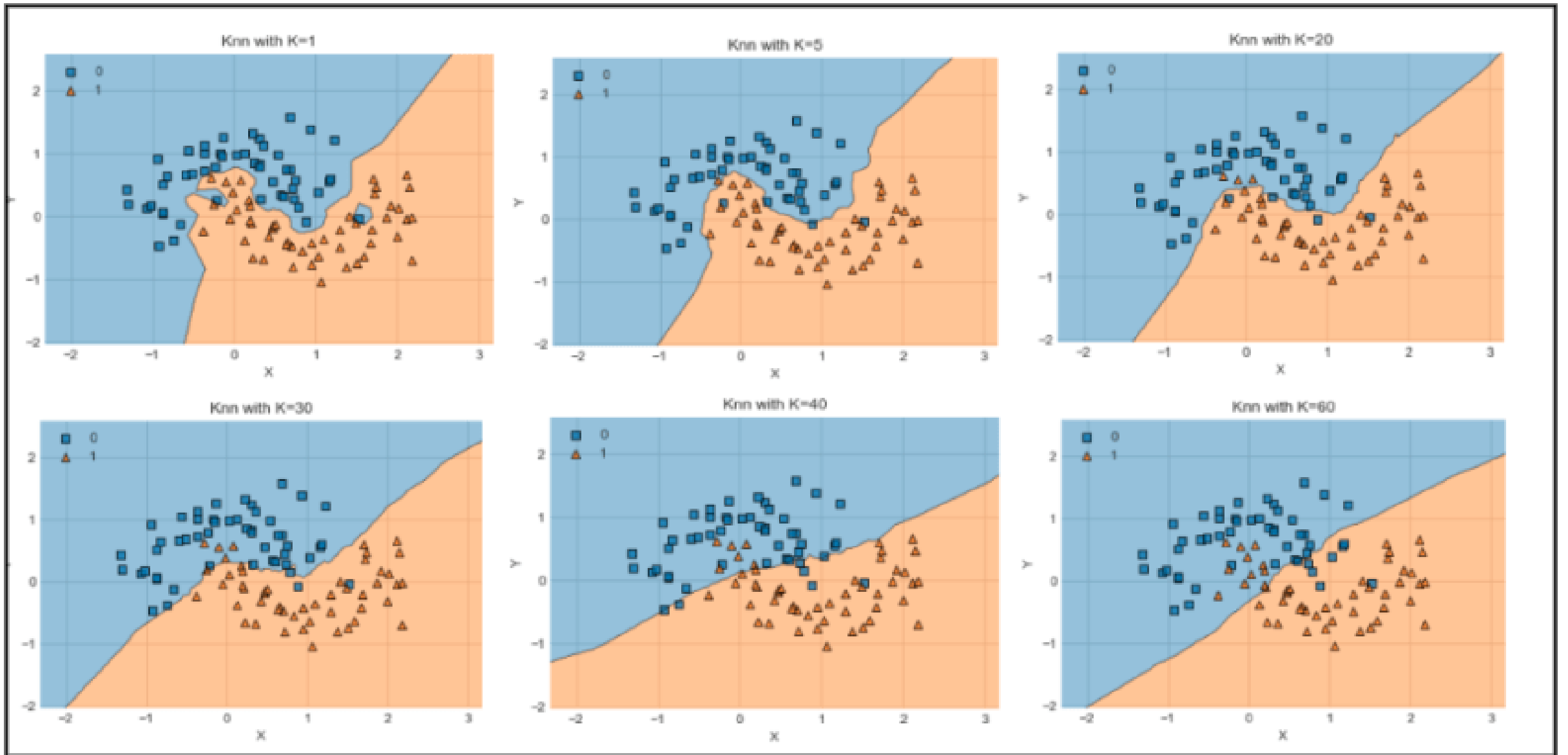
- Punto a clasificar
- Categoría roja
- Categoría azul

Con $k = 1$, el punto blanco se clasifica azul

Con $k = 3$, el punto blanco se clasifica rojo

Hiperparámetro k

- Con k muy grande, hay baja varianza y mucho sesgo (hacia la clase mayor).
- Con k muy bajo, hay mucho ruido, alta varianza, bajo sesgo, hay inestabilidad.
- <https://aprendeia.com/bias-y-varianza-en-machine-learning/>





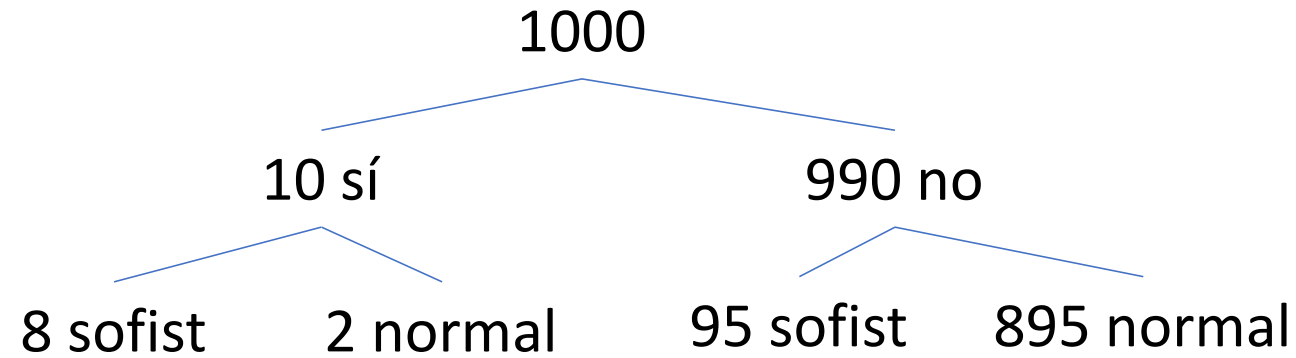
Pontificia Universidad
JAVERIANA
Colombia

Ejemplo

Se hizo una oferta especial a 1000 clientes actuales, de los cuales 10 la tomaron. De los 10 que tomaron la oferta, 8 estaban clasificados, según un estudio de mercados, como clientes *sofisticados*. Por otro lado, de los 990 que no la tomaron, 95 estaban clasificados como *sofisticados*.

Si 10 nuevos clientes son clasificados como *sofisticados*, ¿cuántos cree usted que comprarán el nuevo producto?

Árbol Bayesiano



$$P(\text{sí}/\text{sof}) = \frac{8}{8 + 95}$$

Enunciado

$$P(A/B) = \frac{P(B/A) * P(A)}{P(B)}$$

$$posterior = \frac{likelihood * prior}{evidence}$$

Enunciado en Predicción de Categorías

$$P(C_{sí}/x) = \frac{P(x/C_{sí}) * P(C_{sí})}{P(x/C_{sí}) * P(C_{sí}) + P(x/C_{no}) * P(C_{no})}$$

$$P(C_i/x) = \frac{P(x/C_i) * P(C_i)}{\sum_{i=1}^k P(x/C_i) * P(C_i)}$$

En aprendizaje de máquina

Para datos nuevos

Probabilidad de una categoría dada una característica

En los datos de “aprendizaje”

Probabilidad de característica dada la categoría

Probabilidad de la categoría

$$P(C_i/x) = \frac{P(x/C_i) * P(C_i)}{\sum_{i=1}^k P(x/C_i) * P(C_i)}$$

Probabilidad de la característica

Naive Bayes

Bayes Generalizado

Para datos nuevos

Probabilidad de una categoría dadas características

En los datos de “aprendizaje”

Probabilidad de características dada la categoría

Probabilidad de la categoría

$$P(C_i/x_{1...n}) = \frac{P(x_{1...n}/C_i) * P(C_i)}{\sum_{i=1}^k P(x_{1...n}/C_i) * P(C_i)}$$

Probabilidad de la característica

Dificultad – Propuesta Solución

- **Dificultad**

- No co-ocurrencia de TODAS las características al tiempo con la característica a predecir.
- Intensivo computacionalmente.

- **Propuesta Solución**

- Suponer independencia de las características.
- Por tanto, la probabilidad conjunta es la multiplicación de las marginales.

Naive Bayes

$$P(C_i/x_{1...n}) = \frac{P(x_1/C_i) * P(x_2/C_i) \dots * P(x_n/C_i) * P(C_i)}{\sum_{i=1}^k P(x_1/C_i) * P(x_2/C_i) \dots * P(x_n/C_i) * P(C_i)}$$

$$P(C_i/x_{1...n}) = \frac{\prod_{j=1}^n P(x_j/C_i) * P(C_i)}{\sum_{i=1}^k \prod_{j=1}^n P(x_j/C_i) * P(C_i)}$$

Ejemplo

| Obs | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|-----|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|
| X1 | A | A | A | A | A | B | B | B | B | B | C | C | C | C | C |
| X2 | S | M | M | S | S | S | M | M | L | L | L | M | M | L | L |
| Y | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |

Cuál es la clasificación de un nuevo registro: (B, S) ?

Variantes en Naive Bayes

- $P(x_i/C_i)=0$

Para evitar que algunas probabilidades se vuelvan cero se usa suavizamiento de Laplace con algún valor.

<https://towardsdatascience.com/laplace-smoothing-in-na%C3%AFve-bayes-algorithm-9c237a8bdece>

- **Variables continuas como predictoras**

Se usa una aproximación basada en una función de probabilidad. Usualmente se usa la curva normal.