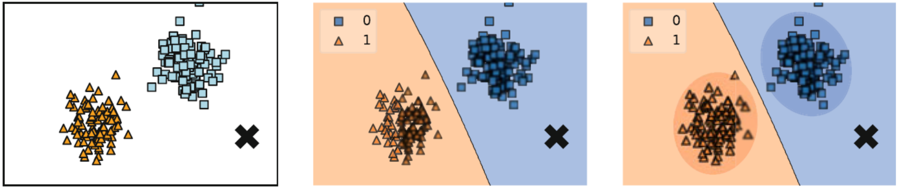
Machine Learning para Inteligencia Artificial

Gaussian Naive Bayes

Universidad ORT Uruguay

11 de Junio, 2025

Modelo Discriminativo	**Modelo Generativo**
El modelo intenta aprender $p(y \mid x)$	El modelo intenta aprender $p(x, y)$ Podemos deducir $p(y \mid x)$ y $p(x \mid y)$



Data | $p(y|\mathbf{x})$ | $p(\mathbf{x}, y) = p(y|\mathbf{x})\ p(\mathbf{x})$

$p(blue|\mathbf{x})$ is high
= certain decision!

$p(blue|\mathbf{x})$ is high
and $p(\mathbf{x})$ is low
= uncertain decision!

Fórmula de Bayes

- En aprendizaje supervisado se usa Bayes de la siguiente forma:

$$p(\text{target} \mid \text{feature}) = \frac{p(\text{feature} \mid \text{target}) \cdot p(\text{target})}{p(\text{feature})}$$

- En clasificación con $\{c_1, \ldots, c_m\}$ clases, concretamente queda:

$$\text{Dado el input } \boldsymbol{x} \rightsquigarrow p(y = c_i \mid \boldsymbol{x}) = \frac{p(\boldsymbol{x} \mid y = c_i) p(y = c_i)}{p(\boldsymbol{x})}$$

- La marginal de $\boldsymbol{x}$ se puede calcular usando probabilidad total:

$$p(\boldsymbol{x}) = \sum_{i=1}^{m} p(\boldsymbol{x} \mid y = c_i) p(y = c_i)$$

Prior - Likelihood - Posterior

Basta considerar las siguiente probabilidades:

- **Class Prior**: $p(y = c_i)$

  Probabilidad a priori de la clase c_i, antes de observar las features $\boldsymbol{x}$.

- **Feature Likelihood**: $p(\boldsymbol{x} \mid y = c_i)$

  Verosimilitud (credibilidad) de las features $\boldsymbol{x}$ dado que la clase es c_i.

- **Class Posterior**: $p(y = c_i \mid \boldsymbol{x})$

  Probabilidad a posteriori de que la clase sea c_i si observamos las features $\boldsymbol{x}$.

Naive Bayes

Dado un conjunto de datos de entrenamiento $T = \{(\boldsymbol{x}, y)\}$

- **Class Prior**: se estima contando

$$p(y = c_i) = \frac{\#\{(\boldsymbol{x}, y) \in T : y = c_i\}}{\# T} = \frac{N_i}{N}$$

- **Naive Assumption**: las features son condicionalmente independientes dada una clase:

$$p(\boldsymbol{x} \mid y = c_i) = p\left(x^{(1)}, \ldots, x^{(D)} \mid y = c_i\right) = \prod_{k=1}^{D} p\left(x^{(k)} \mid y = c_i\right)$$

- Según sea el caso $p\left(x^{(k)} \mid y = c_i\right)$ puede modelarse de diferentes maneras.

Gaussian Naive Bayes

- Se asume que la likelihood de las features es Gaussiana (normal).
- Los parámetros de la Gaussiana para cada clase c_i y se estiman mediante máxima verosimilitud:
  - Media (μ): $\hat{\mu}_{ki} = \frac{1}{N_i} \sum_{(\boldsymbol{x},y):y=c_i} x^{(k)}$
  - Varianza (σ^2): $\hat{\sigma}^2_{ki} = \frac{1}{N_i} \sum_{(\boldsymbol{x},y):y=c_i} \left(x^{(k)} - \hat{\mu}_i\right)^2$
- Recordar que la función de densidad de la Gaussiana es:

$$p(x \mid y = c_i) = \frac{1}{\sqrt{2\pi\sigma^2_{ki}}} \exp\left(-\frac{(x - \mu_{ki})^2}{2\sigma^2_{ki}}\right)$$

donde μ_{ki} y σ^2_{ki} son la media y varianza de la feature k para la clase c_i.