Contrastes de hipótesis. Enfoque de Neyman-Pearson

Ricardo Alberich, Juan Gabriel Gomila y Arnau Mir

Enfoque de Neyman-Pearson

Introducción

- En esta presentación vamos a introducir los contrastes de hipótesis desde una perspectiva más amplia.

Introducción

- En esta presentación vamos a introducir los contrastes de hipótesis desde una perspectiva más amplia.
- Esta nueva perspectiva nos permitirá introducir la clasificación en machine learning y detección en procesamiento de señales.

Introducción

- En esta presentación vamos a introducir los contrastes de hipótesis desde una perspectiva más amplia.
- Esta nueva perspectiva nos permitirá introducir la clasificación en machine learning y detección en procesamiento de señales.
- Este nuevo enfoque se llama enfoque de Neyman-Pearson.

Distribuciones de las hipótesis nula y alternativa

- En los contrastes de hipótesis paramétricos introducidos durante el curso, nos concentramos en la hipótesis nula H_0.

Distribuciones de las hipótesis nula y alternativa

- En los contrastes de hipótesis paramétricos introducidos durante el curso, nos concentramos en la hipótesis nula H_0.
- Tanto si usamos los z-test como los t-test, en el cálculo del p-valor, se usaba la distribución suponiendo que la hipótesis nula H_0 es cierta.

Distribuciones de las hipótesis nula y alternativa

- Para fijar ideas, suponiendo que el contraste considerado era sobre la media μ, y el contraste era de la forma:

$$\left.\begin{array}{ll} H_0 : \mu & = \mu_0, \\ H_1 : \mu & \neq (<,>)\mu_0, \end{array}\right\}$$

donde hemos considerados los tres casos de hipótesis alternativa H_1, suponiendo que la desviación típica σ de la población es conocida, suponemos que la distribución del estadístico de contraste $Z = \frac{\overline{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$ es una normal estándar o $N(0,1)$ suponiendo que la hipótesis nula H_0 es cierta o que $\mu = \mu_0$

Distribuciones de las hipótesis nula y alternativa

- La suposición anterior es equivalente a suponer que la distribución de la media muestral $\overline{X}$ es normal de parámetros $\mu_{\overline{X}} = \mu_0$ y $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$: $\overline{X} = N\left(\mu_0, \frac{\sigma}{\sqrt{n}}\right)$.

Distribuciones de las hipótesis nula y alternativa

- La suposición anterior es equivalente a suponer que la distribución de la media muestral $\overline{X}$ es normal de parámetros $\mu_{\overline{X}} = \mu_0$ y $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$: $\overline{X} = N\left(\mu_0, \frac{\sigma}{\sqrt{n}}\right)$.
- Recordemos que para aceptar o rechazar la hipótesis nula H_0, comparamos el valor crítico $z_\alpha (H_1 : \mu < \mu_0)$, $z_{1-\alpha}(H_1 : \mu > \mu_0)$, $z_{1-\frac{\alpha}{2}}(H_1 : \mu \neq \mu_0)$ con el valor Z del estadístico de contraste y dependiendo de dicha comparación aceptamos o rechazamos la hipótesis nula H_0, siendo α el nivel de significación.

Distribuciones de las hipótesis nula y alternativa

- Concretamente,

Distribuciones de las hipótesis nula y alternativa

- Concretamente,
  - Si $H_1 : \mu < \mu_0$,

Distribuciones de las hipótesis nula y alternativa

- Concretamente,
  - Si $H_1 : \mu < \mu_0$,
    - si $Z < z_\alpha$, rechazamos H_0 y en caso contrario, aceptamos H_0,

Distribuciones de las hipótesis nula y alternativa

- Concretamente,
  - Si $H_1 : \mu < \mu_0$,
    - si $Z < z_\alpha$, rechazamos H_0 y en caso contrario, aceptamos H_0,
  - Si $H_1 : \mu > \mu_0$,

Distribuciones de las hipótesis nula y alternativa

- Concretamente,
  - Si $H_1 : \mu < \mu_0$,
    - si $Z < z_\alpha$, rechazamos H_0 y en caso contrario, aceptamos H_0,
  - Si $H_1 : \mu > \mu_0$,
    - si $Z > z_{1-\alpha}$, rechazamos H_0 y en caso contrario, aceptamos H_0,

Distribuciones de las hipótesis nula y alternativa

- Concretamente,
  - Si $H_1 : \mu < \mu_0$,
    - si $Z < z_\alpha$, rechazamos H_0 y en caso contrario, aceptamos H_0,
  - Si $H_1 : \mu > \mu_0$,
    - si $Z > z_{1-\alpha}$, rechazamos H_0 y en caso contrario, aceptamos H_0,
  - Si $H_1 : \mu \neq \mu_0$,

Distribuciones de las hipótesis nula y alternativa

- Concretamente,
  - Si $H_1 : \mu < \mu_0$,
    - si $Z < z_\alpha$, rechazamos H_0 y en caso contrario, aceptamos H_0,
  - Si $H_1 : \mu > \mu_0$,
    - si $Z > z_{1-\alpha}$, rechazamos H_0 y en caso contrario, aceptamos H_0,
  - Si $H_1 : \mu \neq \mu_0$,
    - si $|Z| > z_{1-\frac{\alpha}{2}}$, rechazamos H_0 y en caso contrario, aceptamos H_0.

Distribuciones de las hipótesis nula y alternativa

- Concretamente,
  - Si $H_1 : \mu < \mu_0$,
    - si $Z < z_\alpha$, rechazamos H_0 y en caso contrario, aceptamos H_0,
  - Si $H_1 : \mu > \mu_0$,
    - si $Z > z_{1-\alpha}$, rechazamos H_0 y en caso contrario, aceptamos H_0,
  - Si $H_1 : \mu \neq \mu_0$,
    - si $|Z| > z_{1-\frac{\alpha}{2}}$, rechazamos H_0 y en caso contrario, aceptamos H_0.
- Un aspecto importante del contrate de hipótesis es la hipótesis alternativa H_1, donde si suponemos que es cierta, la distribución del estadístico de contraste Z o $\overline{X}$ (dependiendo de cuál usemos) sería diferente.

Distribuciones de las hipótesis nula y alternativa

- Concretamente, si la hipótesis alternativa H_1 es cierta, la distribución del estadístico de contraste $\overline{X}$ sería $\overline{X} = N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ o $Z = \frac{\overline{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}} = N\left(\frac{\mu-\mu_0}{\frac{\sigma}{\sqrt{n}}}, 1\right)$.

Distribuciones de las hipótesis nula y alternativa

- Concretamente, si la hipótesis alternativa H_1 es cierta, la distribución del estadístico de contraste $\overline{X}$ sería $\overline{X} = N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ o $Z = \frac{\overline{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}} = N\left(\frac{\mu-\mu_0}{\frac{\sigma}{\sqrt{n}}}, 1\right)$.

- Para usar ambas distribuciones, definimos:

$$f_0(z) = f_Z(z|H_0), \quad f_1(z) = f_Z(z|H_1),$$

las funciones de densidad del estadístico de contraste suponiendo que H_0 es cierta ($f_0(z)$) o suponiendo que H_1 es cierta ($f_1(z)$).

Distribuciones de las hipótesis nula y alternativa

- Concretamente, si la hipótesis alternativa H_1 es cierta, la distribución del estadístico de contraste $\overline{X}$ sería $\overline{X} = N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ o $Z = \frac{\overline{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}} = N\left(\frac{\mu-\mu_0}{\frac{\sigma}{\sqrt{n}}}, 1\right)$.

- Para usar ambas distribuciones, definimos:

$$f_0(z) = f_Z(z|H_0), \quad f_1(z) = f_Z(z|H_1),$$

  las funciones de densidad del estadístico de contraste suponiendo que H_0 es cierta ($f_0(z)$) o suponiendo que H_1 es cierta ($f_1(z)$).

- En el primer caso o suponiendo que H_0 es cierta, $f_0(z)$ sería la función de densidad de una $N(0,1)$ y en el segundo caso o suponiendo que H_1 es cierta, $f_1(z)$ sería la función de densidad de una $N\left(\frac{\mu-\mu_0}{\frac{\sigma}{\sqrt{n}}}, 1\right)$.

Distribuciones de las hipótesis nula y alternativa. Ejemplo

- Suponemos que nos planteamos el contraste de hipótesis siguiente:

$$\left.\begin{aligned} H_0 : \mu &= 2, \\ H_1 : \mu &> 2, \end{aligned}\right\}$$

Distribuciones de las hipótesis nula y alternativa. Ejemplo

- Suponemos que nos planteamos el contraste de hipótesis siguiente:

$$\left.\begin{array}{ll} H_0 : \mu & = 2, \\ H_1 : \mu & > 2, \end{array}\right\}$$

- En el gráfico siguiente hemos dibujado las funciones de densidad f_0 y f_1. La curva azul es f_0 y la roja f_1, donde hemos supuesto que $\mu = 2.5$ y $\sigma = 2$.

Distribuciones de las hipótesis nula y alternativa. Ejemplo

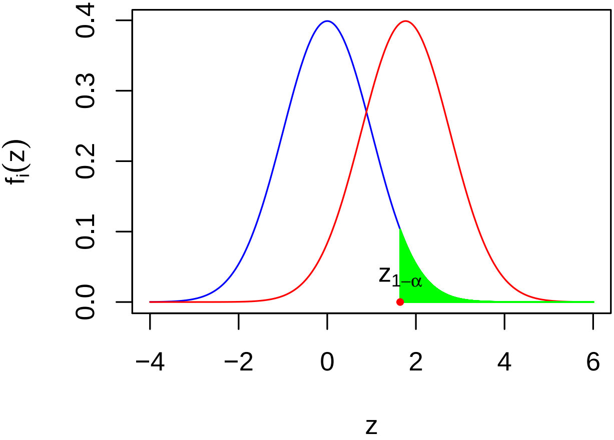
- Suponemos que nos planteamos el contraste de hipótesis siguiente:

$$\left.\begin{array}{ll} H_0 : \mu & = 2, \\ H_1 : \mu & > 2, \end{array}\right\}$$

- En el gráfico siguiente hemos dibujado las funciones de densidad f_0 y f_1. La curva azul es f_0 y la roja f_1, donde hemos supuesto que $\mu = 2.5$ y $\sigma = 2$.

- Rechazaremos la hipótesis nula H_0 si el valor del estadístico de contraste Z se encuentra en la zona verde, o $Z > z_{1-\alpha}$, donde hemos considerado un nivel de significación $\alpha = 0.05$, de donde $z_{0.95} = 1.645$.

Distribuciones de las hipótesis nula y alternativa. Ejemplo



Reglas de decisión

- Un contraste de hipótesis, se basa en una regla de decisión $\delta(\cdot)$ a partir de un espacio de valores $\mathcal{Z}$ del estadístico de contraste Z.

Reglas de decisión

- Un contraste de hipótesis, se basa en una regla de decisión $\delta(\cdot)$ a partir de un espacio de valores $\mathcal{Z}$ del estadístico de contraste Z.

- Concretamente, en un contraste de la media μ del tipo:

$$\left.\begin{array}{ll} H_0 : \mu & = \mu_0, \\ H_1 : \mu & > \mu_0, \end{array}\right\}$$

como la del ejemplo anterior,

Reglas de decisión

- Un contraste de hipótesis, se basa en una regla de decisión $\delta(\cdot)$ a partir de un espacio de valores $\mathcal{Z}$ del estadístico de contraste Z.

- Concretamente, en un contraste de la media μ del tipo:

$$\left.\begin{array}{ll} H_0 : \mu & = \mu_0, \\ H_1 : \mu & > \mu_0, \end{array}\right\}$$

  como la del ejemplo anterior,

  - el espacio de valores $\mathcal{Z}$ sería el conjunto de números reales $\mathbb{R}$,

Reglas de decisión

- Un contraste de hipótesis, se basa en una regla de decisión $\delta(\cdot)$ a partir de un espacio de valores $\mathcal{Z}$ del estadístico de contraste Z.

- Concretamente, en un contraste de la media μ del tipo:

$$\left. \begin{array}{ll} H_0 : \mu & = \mu_0, \\ H_1 : \mu & > \mu_0, \end{array} \right\}$$

  como la del ejemplo anterior,

  - el espacio de valores $\mathcal{Z}$ sería el conjunto de números reales $\mathbb{R}$,
  - el estadístico de contraste Z sería $Z = \frac{\overline{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$ (suponemos σ conocida) y

Reglas de decisión

- Un contraste de hipótesis, se basa en una regla de decisión $\delta(\cdot)$ a partir de un espacio de valores $\mathcal{Z}$ del estadístico de contraste Z.

- Concretamente, en un contraste de la media μ del tipo:

$$\left.\begin{array}{ll} H_0 : \mu & = \mu_0, \\ H_1 : \mu & > \mu_0, \end{array}\right\}$$

  como la del ejemplo anterior,

  - el espacio de valores $\mathcal{Z}$ sería el conjunto de números reales $\mathbb{R}$,
  - el estadístico de contraste Z sería $Z = \frac{\overline{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$ (suponemos σ conocida) y
  - la regla de decisión $\delta(z, \alpha)$ que depende del valor del estadístico de contraste y del nivel de significación α sería:

Reglas de decisión

- Un contraste de hipótesis, se basa en una regla de decisión $\delta(\cdot)$ a partir de un espacio de valores $\mathcal{Z}$ del estadístico de contraste Z.

- Concretamente, en un contraste de la media μ del tipo:

$$\left.\begin{array}{ll} H_0 : \mu & = \mu_0, \\ H_1 : \mu & > \mu_0, \end{array}\right\}$$

  como la del ejemplo anterior,

  - el espacio de valores $\mathcal{Z}$ sería el conjunto de números reales $\mathbb{R}$,
  - el estadístico de contraste Z sería $Z = \frac{\overline{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$ (suponemos σ conocida) y
  - la regla de decisión $\delta(z, \alpha)$ que depende del valor del estadístico de contraste y del nivel de significación α sería:
    - si $Z \geq z_{1-\alpha}$, rechazamos la hipótesis nula H_0,

Reglas de decisión

- Un contraste de hipótesis, se basa en una regla de decisión $\delta(\cdot)$ a partir de un espacio de valores $\mathcal{Z}$ del estadístico de contraste Z.

- Concretamente, en un contraste de la media μ del tipo:

$$\left.\begin{array}{ll} H_0 : \mu & = \mu_0, \\ H_1 : \mu & > \mu_0, \end{array}\right\}$$

como la del ejemplo anterior,

  - el espacio de valores $\mathcal{Z}$ sería el conjunto de números reales $\mathbb{R}$,
  - el estadístico de contraste Z sería $Z = \frac{\overline{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$ (suponemos σ conocida) y
  - la regla de decisión $\delta(z, \alpha)$ que depende del valor del estadístico de contraste y del nivel de significación α sería:
    - si $Z \geq z_{1-\alpha}$, rechazamos la hipótesis nula H_0,
    - si $Z < z_{1-\alpha}$, aceptamos la hipótesis nula H_0.

Reglas de decisión

- Escribiremos la regla de decisión de la forma siguiente:

$$\delta(z, \alpha) = \begin{cases} 1, & \text{si } z \in R_\alpha (\text{ rechazamos } H_0), \\ 0, & \text{si } z \notin R_\alpha (\text{ aceptamos } H_0), \end{cases}$$

donde R_α es la llamada zona de rechazo que valdría en el ejemplo que vamos desarrollando:

$$R_\alpha = \{z \geq z_{1-\alpha} = \phi^{-1}(1-\alpha)\},$$

donde $\phi(z)$ representa la función de distribución de la $N(0,1)$: $\phi(z) = P(Z \leq z)$.

Ejemplo anterior

- En el ejemplo anterior, la regla de decisión para $\alpha = 0.05$ sería:

$$\delta(z, 0.05) = \begin{cases} 1, & \text{si } z \in R_{0.05}(\text{ rechazamos } H_0), \\ 0, & \text{si } z \notin R_{0.05}(\text{ aceptamos } H_0), \end{cases}$$

donde $R_{0.05} = \{z \geq z_{0.95} = \phi^{-1}(0.95) = 1.645\}$.

Errores tipo I y tipo II

- Recordemos que en un contraste de hipótesis, el error tipo I se definía como:

$$\text{Error tipo I} = P(\text{Rechazar } H_0 | H0 \text{ cierta }),$$

o como la probabilidad de rechazar la hipótesis nula suponiendo ésta cierta.

Errores tipo I y tipo II

- Recordemos que en un contraste de hipótesis, el error tipo I se definía como:

$$\text{Error tipo I} = P(\text{Rechazar } H_0 | H0 \text{ cierta }),$$

  o como la probabilidad de rechazar la hipótesis nula suponiendo ésta cierta.

- De la misma manera, el error tipo II se definía como:

$$\text{Error tipo II} = P(\text{Aceptar } H_0 | H0 \text{ falsa }),$$

  o como la probabilidad de aceptar la hipótesis nula suponiendo ésta falsa.

Errores tipo I y tipo II

- Si interpretamos un contraste de hipótesis como un proceso de decisión,

Errores tipo I y tipo II

- Si interpretamos un contraste de hipótesis como un proceso de decisión,
  - aceptar la hipótesis nula sería tener un valor negativo y

Errores tipo I y tipo II

- Si interpretamos un contraste de hipótesis como un proceso de decisión,
  - aceptar la hipótesis nula sería tener un valor negativo y
  - rechazarla sería tener un valor positivo.

Errores tipo I y tipo II

- Si interpretamos un contraste de hipótesis como un proceso de decisión,
  - aceptar la hipótesis nula sería tener un valor negativo y
  - rechazarla sería tener un valor positivo.
- Entonces, podemos interpretar

Errores tipo I y tipo II

- Si interpretamos un contraste de hipótesis como un proceso de decisión,
  - aceptar la hipótesis nula sería tener un valor negativo y
  - rechazarla sería tener un valor positivo.
- Entonces, podemos interpretar
  - el error tipo I como un falso positivo ya que declaramos positiva una decisión que debería ser negativa y

Errores tipo I y tipo II

- Si interpretamos un contraste de hipótesis como un proceso de decisión,
    - aceptar la hipótesis nula sería tener un valor negativo y
    - rechazarla sería tener un valor positivo.
- Entonces, podemos interpretar
    - el error tipo I como un falso positivo ya que declaramos positiva una decisión que debería ser negativa y
    - y el error tipo II como un falso negativo ya que declaramos positiva una decisión que debería ser positiva.

Tasa de falsos positivos y falsos negativos

- En un contexto del proceso de decisión,

Tasa de falsos positivos y falsos negativos

- En un contexto del proceso de decisión,
  - el error tipo I se podría interpreta como la tasa de falsos positivos y

Tasa de falsos positivos y falsos negativos

- En un contexto del proceso de decisión,
  - el error tipo I se podría interpreta como la tasa de falsos positivos y
  - el error tipo II se podría interpreta como la tasa de falsos negativos.

Tasa de falsos positivos y falsos negativos

- En un contexto del proceso de decisión,
  - el error tipo I se podría interpreta como la tasa de falsos positivos y
  - el error tipo II se podría interpreta como la tasa de falsos negativos.
- Otro concepto que se introdujo en los contrastes de hipótesis es la potencia de un contraste que se definía como:

$$\begin{aligned} \text{Potencia de un contraste} &= 1 - \text{Error tipo II} \\ &= 1 - P(\text{Aceptar } H_0 | H0 \text{ falsa }) \\ &= P(\text{rechazar } H_0 | H_0 \text{ falsa}), \end{aligned}$$

es decir, como la probabilidad de rechazar la hipótesis nula suponiendo ésta falsa.

Tasa de verdaderos positivos

- En un contexto del proceso de decisión, la potencia del contraste se puede interpretar como la probabilidad de detectar un negativo.

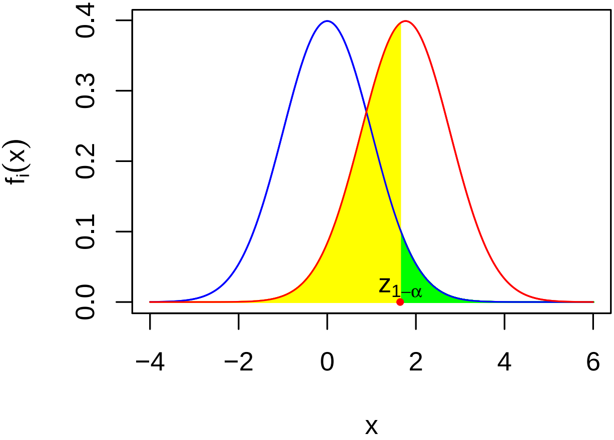
Tasa de verdaderos positivos

- En un contexto del proceso de decisión, la potencia del contraste se puede interpretar como la probabilidad de detectar un negativo.
- Por tanto, la potencia de un contraste puede interpretarse como la tasa de verdaderos positivos o la probabilidad de detección.

Tasa de verdaderos positivos

- En un contexto del proceso de decisión, la potencia del contraste se puede interpretar como la probabilidad de detectar un negativo.
- Por tanto, la potencia de un contraste puede interpretarse como la tasa de verdaderos positivos o la probabilidad de detección.
- En el gráfico siguiente mostramos en amarillo el error tipo II o la tasa de falsos negativos para el ejemplo anterior. Recordemos que en verde está el error tipo I o la tasa de falsos positivos.

Tasa de verdaderos positivos



Tasa de falsos positivos, falsos negativos y verdaderos positivos

- Matemáticamente, podemos escribir las tasas anteriores de la forma siguiente:

Tasa de falsos positivos, falsos negativos y verdaderos positivos

- Matemáticamente, podemos escribir las tasas anteriores de la forma siguiente:
  - Tasa de falsos positivos o error tipo I:

$$P_{FP} = \int_{R_\alpha} f_0(z)\, dz = \int_{z_{1-\alpha}}^{\infty} f_0(z)\, dz.$$

Tasa de falsos positivos, falsos negativos y verdaderos positivos

- Matemáticamente, podemos escribir las tasas anteriores de la forma siguiente:
  - Tasa de falsos positivos o error tipo I:

$$P_{FP} = \int_{R_\alpha} f_0(z)\,dz = \int_{z_{1-\alpha}}^{\infty} f_0(z)\,dz.$$

  - Tasa de falsos negativos o error tipo II:

$$P_{FN} = \int_{R_\alpha^c} f_1(z)\,dz = \int_{-\infty}^{z_{1-\alpha}} f_1(z)\,dz.$$

Tasa de falsos positivos, falsos negativos y verdaderos positivos

- Matemáticamente, podemos escribir las tasas anteriores de la forma siguiente:
  - Tasa de falsos positivos o error tipo I:

$$P_{FP} = \int_{R_\alpha} f_0(z)\,dz = \int_{z_{1-\alpha}}^{\infty} f_0(z)\,dz.$$

  - Tasa de falsos negativos o error tipo II:

$$P_{FN} = \int_{R_\alpha^c} f_1(z)\,dz = \int_{-\infty}^{z_{1-\alpha}} f_1(z)\,dz.$$

  - Tasa de verdaderos positivos o potencia del contraste:

$$P_{VP} = \int_{R_\alpha} f_1(z)\,dz = \int_{z_{1-\alpha}}^{\infty} f_1(z)\,dz = 1 - P_{FN}.$$

Tasa de falsos positivos, falsos negativos y verdaderos positivos

- Podemos escribir las probabilidades anteriores en función de la regla de decisión $\delta(z, \alpha)$ de la forma siguiente:

$$
\begin{aligned}
P_{FP} &= \int_{\mathbb{R}} \delta(z, \alpha) f_0(z)\, dz = \int_{z_{1-\alpha}}^{\infty} f_0(z)\, dz, \\
P_{FN} &= \int_{\mathbb{R}} (1 - \delta(z, \alpha)) f_1(z)\, dz = \int_{-\infty}^{z_{1-\alpha}} f_1(z)\, dz, \\
P_{VP} &= \int_{\mathbb{R}} \delta(z, \alpha) f_1(z)\, dz = \int_{z_{1-\alpha}}^{\infty} f_1(z)\, dz.
\end{aligned}
$$

Ejemplo anterior

- En el ejemplo anterior, las tasas de falsos positivos, falsos negativos y verdaderos positivos son las siguientes:

$$P_{FP} = \int_{1.645}^{\infty} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{z^2}{2}}\, dz = 1 - \phi(1.645) = 0.05,$$

$$P_{FN} = \int_{-\infty}^{1.645} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{\left(z-\frac{2.5-2}{\frac{2}{\sqrt{50}}}\right)^2}{2}}\, dz = \int_{-\infty}^{1.645} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{(z-1.768)}{2}}\, dz$$

$$= P(N(1.768, 1) \leq 1.645) = \phi(1.645 - 1.768) = \phi(-0.123)$$

$$= 0.451,$$

$$P_{VP} = 1 - P_{FN} = 1 - 0.451 = 0.549.$$

Decisión de Neyman-Pearson

- Una vez establecidos los conceptos necesarios, el objetivo es hallar la mejor regla de decisión para un contraste de hipótesis.

Definición

La regla de decisión de Neyman-Pearson se define como el siguiente problema de optimización:

$$\hat{\delta} = \max_{\delta \text{ tal que } P_{FP}(\delta) \leq \alpha} P_{VP}(\delta)$$

Decisión de Neyman-Pearson

- Una vez establecidos los conceptos necesarios, el objetivo es hallar la mejor regla de decisión para un contraste de hipótesis.
- Definimos la regla de decisión de Neyman-Pearson precisamente como la mejor regla de decisión donde formalmente se define como el siguiente problema de optimización:

Definición

La regla de decisión de Neyman-Pearson se define como el siguiente problema de optimización:

$$\hat{\delta} = \max_{\delta \text{ tal que } P_{FP}(\delta) \leq \alpha} P_{VP}(\delta)$$

Decisión de Neyman-Pearson

- Es decir, de entre todas las reglas de decisión δ tal que la tasa de falsos positivos es menor que un cierto nivel de significación α, hallar la que maximiza la tasa de verdaderos positivos o la tasa de detección.

Decisión de Neyman-Pearson

- Es decir, de entre todas las reglas de decisión δ tal que la tasa de falsos positivos es menor que un cierto nivel de significación α, hallar la que maximiza la tasa de verdaderos positivos o la tasa de detección.

- Para resolver el problema de optimización anterior, necesitamos introducir el índice de verosimilitud entre dos distribuciones de funciones de densidad $f_0(z)$ y $f_1(z)$ como:

$$L(z) = \frac{f_1(z)}{f_0(z)}.$$

Decisión de Neyman-Pearson

- El siguiente resultado resuelve el problema de optimización:

Decisión de Neyman-Pearson

- El siguiente resultado resuelve el problema de optimización:

Teorema

La solución al problema de optimización de la regla de decisión de Neyman-Pearson es el siguiente:

$$\hat{\delta}(z) = \begin{cases} 1, & \text{si } L(z) \geq \eta, \\ 0, & \text{si } L(z) < \eta, \end{cases}$$

donde η depende del nivel de significación α.

Decisión de Neyman-Pearson

- El siguiente resultado resuelve el problema de optimización:

Teorema

La solución al problema de optimización de la regla de decisión de Neyman-Pearson es el siguiente:

$$\hat{\delta}(z) = \begin{cases} 1, & \textit{si } L(z) \geq \eta, \\ 0, & \textit{si } L(z) < \eta, \end{cases}$$

donde η depende del nivel de significación α.

- El teorema anterior dice que si el objetivo es maximizar la tasa de detección o la tasa de verdaderos positivos manteniendo la tasa de falsos positivos, no podemos hacerlo mejor que la regla de decisión dada por el Teorema.

Demostración del Teorema

- La relación entre η y α es la siguiente:

$$\alpha = P_{FP}(\hat{\delta}) = \int_{\mathbb{R}} \hat{\delta}(z) f_0(z)\, dz = \int_{L(z) \geq \eta} f_0(z)\, dz$$

Demostración del Teorema

- La relación entre η y α es la siguiente:

$$\alpha = P_{FP}(\hat{\delta}) = \int_{\mathbb{R}} \hat{\delta}(z) f_0(z)\, dz = \int_{L(z) \geq \eta} f_0(z)\, dz$$

- Sea δ otra regla de decisión. Nuestro objetivo es demostrar que $P_{VP}(\hat{\delta}) \geq P_{VP}(\delta)$.

Demostración del Teorema

- La relación entre η y α es la siguiente:

$$\alpha = P_{FP}(\hat{\delta}) = \int_{\mathbb{R}} \hat{\delta}(z) f_0(z)\, dz = \int_{L(z)\geq\eta} f_0(z)\, dz$$

- Sea δ otra regla de decisión. Nuestro objetivo es demostrar que $P_{VP}(\hat{\delta}) \geq P_{VP}(\delta)$.

- Como la regla δ debe cumplir que la tasa de falsos positivos debe ser menor que α, tenemos que:

$$\begin{aligned} \alpha \geq P_{FP}(\delta) &= \int_{\mathbb{R}} \delta(z) f_0(z)\, dz \\ &= \int_{L(z)\geq\eta} \delta(z) f_0(z)\, dz + \int_{L(z)<\eta} \delta(z) f_0(z)\, dz. \end{aligned}$$

Demostración del Teorema (continuación)

- Entonces:
$$\int_{L(z)\geq\eta} f_0(z)\,dz \geq \int_{L(z)\geq\eta} \delta(z) f_0(z)\,dz + \int_{L(z)<\eta} \delta(z) f_0(z)\,dz, \;\Rightarrow$$
$$\int_{L(z)\geq\eta} (1-\delta(z)) f_0(z)\,dz - \int_{L(z)<\eta} \delta(z) f_0(z)\,dz \geq 0.$$

Demostración del Teorema (continuación)

- Entonces:

$$\int_{L(z)\geq\eta} f_0(z)\,dz \geq \int_{L(z)\geq\eta} \delta(z) f_0(z)\,dz + \int_{L(z)<\eta} \delta(z) f_0(z)\,dz, \Rightarrow$$

$$\int_{L(z)\geq\eta} (1-\delta(z)) f_0(z)\,dz - \int_{L(z)<\eta} \delta(z) f_0(z)\,dz \geq 0.$$

- A continuación, veamos que $P_{VP}(\hat{\delta}) \geq P_{VP}(\delta)$:

$$P_{VP}(\hat{\delta}) = \int_{\mathbb{R}} \hat{\delta}(z) f_1(z)\,dz = \int_{L(z)\geq\eta} f_1(z)\,dz,$$

$$P_{VP}(\delta) = \int_{\mathbb{R}} \delta(z) f_1(z)\,dz = \int_{L(z)\geq\eta} \delta(z) f_1(z)\,dz$$

$$+ \int_{L(z)<\eta} \delta(z) f_1(z)\,dz,$$

Demostración del Teorema (continuación)

- Restando las dos expresiones anteriores,

$$P_{VP}(\hat{\delta}) - P_{VP}(\delta) = \int_{L(z)\geq\eta} (1-\delta(z)) f_1(z)\, dz - \int_{L(z)<\eta} \delta(z) f_1(z)\, dz$$

Demostración del Teorema (continuación)

- Restando las dos expresiones anteriores,

$$P_{VP}(\hat{\delta})-P_{VP}(\delta) = \int_{L(z)\geq\eta}(1-\delta(z))f_1(z)\,dz-\int_{L(z)<\eta}\delta(z)f_1(z)\,dz$$

- Ahora bien,

Demostración del Teorema (continuación)

- Restando las dos expresiones anteriores,

$$P_{VP}(\hat{\delta})-P_{VP}(\delta) = \int_{L(z)\geq\eta}(1-\delta(z))f_1(z)\,dz-\int_{L(z)<\eta}\delta(z)f_1(z)\,dz$$

- Ahora bien,
  - si $L(z) = \frac{f_1(z)}{f_0(z)} \geq \eta$, $f_1(z) \geq \eta f_0(z)$,

Demostración del Teorema (continuación)

- Restando las dos expresiones anteriores,

$$P_{VP}(\hat{\delta}) - P_{VP}(\delta) = \int_{L(z)\geq\eta} (1-\delta(z)) f_1(z)\, dz - \int_{L(z)<\eta} \delta(z) f_1(z)\, dz$$

- Ahora bien,
  - si $L(z) = \frac{f_1(z)}{f_0(z)} \geq \eta$, $f_1(z) \geq \eta f_0(z)$,
  - si $L(z) = \frac{f_1(z)}{f_0(z)} < \eta$, $f_1(z) < \eta f_0(z)$ y por tanto, $-f_1(z) > -\eta f_0(z)$.

Demostración del Teorema (continuación)

- Entonces:

$$
\begin{aligned}
P_{VP}(\hat{\delta}) - P_{VP}(\delta) &\geq \int_{L(z)\geq\eta} (1-\delta(z))\eta f_0(z)\,dz \\
&\quad - \int_{L(z)<\eta} \eta\delta(z) f_0(z)\,dz \\
&= \eta\left(\int_{L(z)\geq\eta} (1-\delta(z)) f_0(z)\,dz - \int_{L(z)<\eta} \delta(z) f_0(z)\,dz\right)
\end{aligned}
$$

Demostración del Teorema (continuación)

- Entonces:

$$\begin{aligned} P_{VP}(\hat{\delta}) - P_{VP}(\delta) &\geq \int_{L(z)\geq\eta} (1-\delta(z))\eta f_0(z)\,dz \\ &\quad - \int_{L(z)<\eta} \eta\delta(z) f_0(z)\,dz \\ &= \eta\left(\int_{L(z)\geq\eta}(1-\delta(z))f_0(z)\,dz - \int_{L(z)<\eta}\delta(z)f_0(z)\,dz\right) \end{aligned}$$

- Anteriormente vimos que

$$\int_{L(z)\geq\eta}(1-\delta(z))f_0(z)\,dz - \int_{L(z)<\eta}\delta(z)f_0(z)\,dz \geq 0,$$

por tanto,

$$P_{VP}(\hat{\delta}) - P_{VP}(\delta) \geq 0,$$

tal como queríamos demostrar.

Decisión de Neyman-Pearson. Ejemplo

- Ahora ya sabemos cómo hallar reglas de decisión óptimas dado un contraste de hipótesis.

Decisión de Neyman-Pearson. Ejemplo

- Ahora ya sabemos cómo hallar reglas de decisión óptimas dado un contraste de hipótesis.
- Apliquemos el Teorema anterior al ejemplo desarrollado pero lo haremos con cualquier valor de σ.

Decisión de Neyman-Pearson. Ejemplo

- Ahora ya sabemos cómo hallar reglas de decisión óptimas dado un contraste de hipótesis.
- Apliquemos el Teorema anterior al ejemplo desarrollado pero lo haremos con cualquier valor de σ.
- En el ejemplo,

$$f_0(z) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{z^2}{2}}, \quad f_1(z) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{\frac{-\left(z - \frac{\mu - 2}{\frac{\sigma}{\sqrt{n}}}\right)^2}{2}}$$

Decisión de Neyman-Pearson. Ejemplo

- Ahora ya sabemos cómo hallar reglas de decisión óptimas dado un contraste de hipótesis.
- Apliquemos el Teorema anterior al ejemplo desarrollado pero lo haremos con cualquier valor de σ.
- En el ejemplo,

$$f_0(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \quad f_1(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{\left(z - \frac{\mu-2}{\frac{\sigma}{\sqrt{n}}}\right)^2}{2}}$$

- Sea $\mu_1 = \frac{\mu-2}{\frac{\sigma}{\sqrt{n}}}$. El valor de $L(z)$ será:

$$L(z) = \frac{f_1(z)}{f_0(z)} = e^{-\frac{(z-\mu_1)^2}{2} + \frac{z^2}{2}} = e^{\frac{1}{2}(2z\mu_1 - \mu_1^2)}.$$

Decisión de Neyman-Pearson. Ejemplo

- La condición $L(z) \geq \eta$ será en nuestro caso,

$$\mathrm{e}^{\frac{1}{2}(2z\mu_1 - \mu_1^2)} \geq \eta, \Leftrightarrow 2z\mu_1 \geq 2\ln(\eta) + \mu_1^2, \Leftrightarrow z \geq \frac{2\ln(\eta) + \mu_1^2}{2\mu_1} := \tau.$$

Decisión de Neyman-Pearson. Ejemplo

- La condición $L(z) \geq \eta$ será en nuestro caso,

$$\mathrm{e}^{\frac{1}{2}(2z\mu_1-\mu_1^2)} \geq \eta, \Leftrightarrow 2z\mu_1 \geq 2\ln(\eta)+\mu_1^2, \Leftrightarrow z \geq \frac{2\ln(\eta)+\mu_1^2}{2\mu_1} := \tau.$$

- Para hallar τ en función de α, hay que tener en cuenta que:

$$\alpha = P_{FP}(\hat{\delta}) = \int_{L(z)\geq\eta} f_0(z)\,dz = \int_{\tau}^{\infty} \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{z^2}{2}}\,dz = 1-\phi(\tau).$$

Decisión de Neyman-Pearson. Ejemplo

- La condición $L(z) \geq \eta$ será en nuestro caso,

$$\mathrm{e}^{\frac{1}{2}(2z\mu_1-\mu_1^2)} \geq \eta, \Leftrightarrow 2z\mu_1 \geq 2\ln(\eta)+\mu_1^2, \Leftrightarrow z \geq \frac{2\ln(\eta)+\mu_1^2}{2\mu_1} := \tau.$$

- Para hallar τ en función de α, hay que tener en cuenta que:

$$\alpha = P_{FP}(\hat{\delta}) = \int_{L(z)\geq\eta} f_0(z)\,dz = \int_{\tau}^{\infty} \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{z^2}{2}}\,dz = 1-\phi(\tau).$$

- Entonces,

$$\phi(\tau) = 1-\alpha, \Leftrightarrow \tau = \phi^{-1}(1-\alpha).$$

Decisión de Neyman-Pearson. Ejemplo

- En pocas palabras la regla de decisión óptima sería

Decisión de Neyman-Pearson. Ejemplo

- En pocas palabras la regla de decisión óptima sería
  - si $z \geq \phi^{-1}(1-\alpha) = z_{1-\alpha}$, rechazamos la hipótesis nula y,

Decisión de Neyman-Pearson. Ejemplo

- En pocas palabras la regla de decisión óptima sería
  - si $z \geq \phi^{-1}(1-\alpha) = z_{1-\alpha}$, rechazamos la hipótesis nula y,
  - si $z < \phi^{-1}(1-\alpha) = z_{1-\alpha}$, aceptamos la hipótesis nula.

Decisión de Neyman-Pearson. Ejemplo

- En pocas palabras la regla de decisión óptima sería
  - si $z \geq \phi^{-1}(1-\alpha) = z_{1-\alpha}$, rechazamos la hipótesis nula y,
  - si $z < \phi^{-1}(1-\alpha) = z_{1-\alpha}$, aceptamos la hipótesis nula.
- La regla anterior es la regla que aprendimos. En este nuevo enfoque, hemos demostrado cuál es la razón que sea la regla de decisión óptima.

Curva ROC

Introducción

- Como ya hemos comentado, realizar un contraste de hipótesis es similar a realizar una clasificación en dos clases.

Introducción

- Como ya hemos comentado, realizar un contraste de hipótesis es similar a realizar una clasificación en dos clases.
- Rechazar la hipótesis nula sería equivalente a clasificar como positivo y aceptar la hipótesis nula, como negativo.

Introducción

- Como ya hemos comentado, realizar un contraste de hipótesis es similar a realizar una clasificación en dos clases.
- Rechazar la hipótesis nula sería equivalente a clasificar como positivo y aceptar la hipótesis nula, como negativo.
- Entonces, cualquier métrica de evaluación para contrastes de hipótesis tiene su equivalencia en clasificación y viceversa.

Introducción

- Como ya hemos comentado, realizar un contraste de hipótesis es similar a realizar una clasificación en dos clases.
- Rechazar la hipótesis nula sería equivalente a clasificar como positivo y aceptar la hipótesis nula, como negativo.
- Entonces, cualquier métrica de evaluación para contrastes de hipótesis tiene su equivalencia en clasificación y viceversa.
- Vamos a introducir la curva *Receiving Operating Characteristic* o curva ROC. Dicha curva es una de las métricas más usadas en tareas de clasificación, detección y segmentación en machine learning.

Introducción

- Vamos a relacionar el enfoque de Neyman Pearson para contrastes de hipótesis y las curvas ROC.

Introducción

- Vamos a relacionar el enfoque de Neyman Pearson para contrastes de hipótesis y las curvas ROC.
- Recordemos que el criterio de Neyman-Pearson es la regla de decisión óptima $\hat{\delta}$ que resuelve el problema de optimización siguiente:

$$\hat{\delta} = \max_{\delta \text{ tal que } P_{FP}(\delta) \leq \alpha} P_{VP}(\delta)$$

Introducción

- Vamos a relacionar el enfoque de Neyman Pearson para contrastes de hipótesis y las curvas ROC.

- Recordemos que el criterio de Neyman-Pearson es la regla de decisión óptima $\hat{\delta}$ que resuelve el problema de optimización siguiente:

$$\hat{\delta} = \max_{\delta \text{ tal que } P_{FP}(\delta) \leq \alpha} P_{VP}(\delta)$$

- La regla de decisión óptima $\hat{\delta}$ cambia si cambiamos el nivel de significación α. Por tanto, podemos escribir $\hat{\delta}(\alpha)$.

Curva ROC

- Entonces, dados un conjunto de niveles de significación $\alpha_1, \ldots, \alpha_N$, podemos calcular las reglas de decisión óptimas $\hat{\delta}_1, \ldots, \hat{\delta}_N$ correspondientes a dichos niveles de significación.

Curva ROC

- Entonces, dados un conjunto de niveles de significación $\alpha_1, \ldots, \alpha_N$, podemos calcular las reglas de decisión óptimas $\hat{\delta}_1, \ldots, \hat{\delta}_N$ correspondientes a dichos niveles de significación.
- Además, para cada regla de decisión óptima $\hat{\delta}_i$, podemos calcular la tasa de falsos positivos $P_{FP}(\hat{\delta}_i)$ y la tasa de verdaderos positivos $P_{VP}(\hat{\delta}_i)$, $i = 1, \ldots, N$.

Curva ROC

- Entonces, dados un conjunto de niveles de significación $\alpha_1, \ldots, \alpha_N$, podemos calcular las reglas de decisión óptimas $\hat{\delta}_1, \ldots, \hat{\delta}_N$ correspondientes a dichos niveles de significación.
- Además, para cada regla de decisión óptima $\hat{\delta}_i$, podemos calcular la tasa de falsos positivos $P_{FP}(\hat{\delta}_i)$ y la tasa de verdaderos positivos $P_{VP}(\hat{\delta}_i)$, $i = 1, \ldots, N$.
- El gráfico de las tasas de verdaderos positivos $P_{VP}(\hat{\delta}_i)$ (eje Y) como función de las tasas de falsos positivos $P_{FP}(\hat{\delta}_i)$ (eje X), $i = 1, \ldots, N$ se denomina curva ROC para el contraste de hipótesis considerado.

Curva ROC. Ejemplo

- Para el ejemplo considerado, dado un valor de $\alpha \in [0,1]$, la tasa de falsos positivos $P_{FP}(\hat{\delta})$ y la tasa de verdaderos positivos $P_{VP}(\hat{\delta})$ para dicho valor de α serán:

$$P_{FP} = \int_{\phi^{-1}(1-\alpha)}^{\infty} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{z^2}{2}} \, dz = 1 - \phi(\phi^{-1}(1-\alpha)) = \alpha,$$

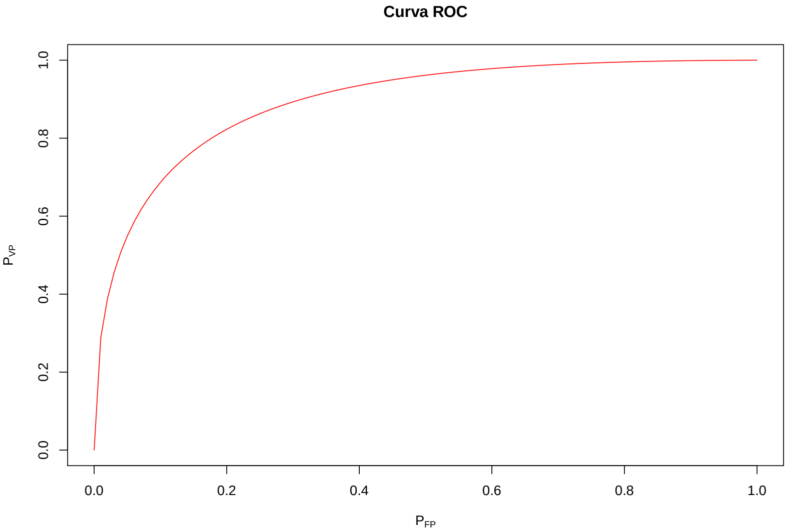
$$\begin{aligned} P_{FN} &= \int_{-\infty}^{\phi^{-1}(1-\alpha)} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{\left(z - \frac{\frac{2.5-2}{2}}{\sqrt{50}}\right)^2}{2}} \, dz \\ &= \int_{-\infty}^{\phi^{-1}(1-\alpha)} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{(z-1.768)}{2}} \, dz \\ &= P(N(1.768, 1) \leq \phi^{-1}(1-\alpha)) = \phi(\phi^{-1}(1-\alpha) - 1.768), \end{aligned}$$

$$P_{VP} = 1 - P_{FN} = 1 - \phi(\phi^{-1}(1-\alpha) - 1.768).$$

Curva ROC. Ejemplo

- El gráfico siguiente muestra la curva ROC para el ejemplo considerado.

Curva ROC. Ejemplo



Interpretación de la curva ROC

- El valor $(0,0)$ siempre está en la curva ROC. Significa que se tiene una tasa nula de falsos positivos y verdaderos positivos. En este caso, siempre aceptamos la hipótesis nula pero tampoco se detecta nada ya que nunca aceptamos la hipótesis alternativa H_1. Estaríamos en un caso sin ningún tipo de interés desde el punto de vista de realizar un contraste de hipótesis o de realizar una clasificación en dos clases.

Interpretación de la curva ROC

- El valor $(0, 0)$ siempre está en la curva ROC. Significa que se tiene una tasa nula de falsos positivos y verdaderos positivos. En este caso, siempre aceptamos la hipótesis nula pero tampoco se detecta nada ya que nunca aceptamos la hipótesis alternativa H_1. Estaríamos en un caso sin ningún tipo de interés desde el punto de vista de realizar un contraste de hipótesis o de realizar una clasificación en dos clases.

- El valor $(1, 1)$ también pertenece siempre a la curva ROC. Significa que se tiene un 100 % de falsos positivos y un 100 % de verdaderos positivos. En este caso, siempre rechazamos la hipótesis nula, por tanto, tenemos un 100 % de detección pero al mismo tiempo tenemos una tasa del 100 % de falsos positivos o un error tipo I máximo.

Interpretación de la curva ROC

- Para cualquier valor de α diferente de 0 y 1, no podemos tener una tasa de verdaderos positivos más alta que la que nos da la curva ROC para la tasa de falsos positivos fijada.

Interpretación de la curva ROC

- Para cualquier valor de α diferente de 0 y 1, no podemos tener una tasa de verdaderos positivos más alta que la que nos da la curva ROC para la tasa de falsos positivos fijada.
- La curva ROC dice cómo cambia la regla de decisión a medida que cambiamos el umbral. Dicho umbral es un concepto usado tanto en contraste de hipótesis como en clasificación.

Interpretación de la curva ROC

- Para cualquier valor de α diferente de 0 y 1, no podemos tener una tasa de verdaderos positivos más alta que la que nos da la curva ROC para la tasa de falsos positivos fijada.
- La curva ROC dice cómo cambia la regla de decisión a medida que cambiamos el umbral. Dicho umbral es un concepto usado tanto en contraste de hipótesis como en clasificación.
- En contraste de hipótesis, el umbral sería el nivel de significación α. En clasificación, el umbral sería un valor para decidir si un valor de la muestra se clasifica como clase 1 o clase 0.

Interpretación de la curva ROC

- La curva ROC ideal sería aquella en la cual la tasa de falsos positivos siempre es nula, $P_{FP} = 0$ y la tasa de verdaderos positivos siempre vale 1 o su valor máximo, $P_{VP} = 1$. En el gráfico, la curva correspondería a los lados del cuadrado que van de $(0, 0)$ a $(0, 1)$ y de $(0, 1)$ a $(1, 1)$. Sin embargo, esta situación es teórica y nunca se da en la práctica.

Decisión ciega

- Imaginemos que para realizar un contraste de hipótesis determinado, consideramos la regla de decisión que consiste en rechazar la hipótesis nula H_0 con probabilidad α y aceptarla con probabilidad $1-\alpha$. Es decir:

$$\delta(z) = \begin{cases} 1, & \text{con probabilidad } \alpha, \\ 0, & \text{con probabilidad } 1-\alpha. \end{cases}$$

  Llamamos a dicha regla de decisión, decisión ciega ya que no depende del valor del estadístico de contraste.

Decisión ciega

- En este caso, fijado un valor del nivel de significación α, la tasa de falsos positivos y la tasa de verdaderos positivos serán

$$P_{FP} = P(\text{ rechazar } H0|H0 \text{ cierta}) = \alpha,$$
$$P_{VP} = P(\text{ rechazar } H0|H0 \text{ cierta}) = \alpha.$$

Es decir el punto de la curva ROC de la decisión ciega sería (α, α).

Decisión ciega

- En este caso, fijado un valor del nivel de significación α, la tasa de falsos positivos y la tasa de verdaderos positivos serán

$$P_{FP} = P(\text{ rechazar } H0|H0 \text{ cierta}) = \alpha,$$
$$P_{VP} = P(\text{ rechazar } H0|H0 \text{ cierta}) = \alpha.$$

  Es decir el punto de la curva ROC de la decisión ciega sería (α, α).

- Por tanto, la curva ROC correspondiente a la decisión ciega sería la diagonal del cuadrado $[0,1] \times [0,1]$.

Decisión ciega

- En este caso, fijado un valor del nivel de significación α, la tasa de falsos positivos y la tasa de verdaderos positivos serán

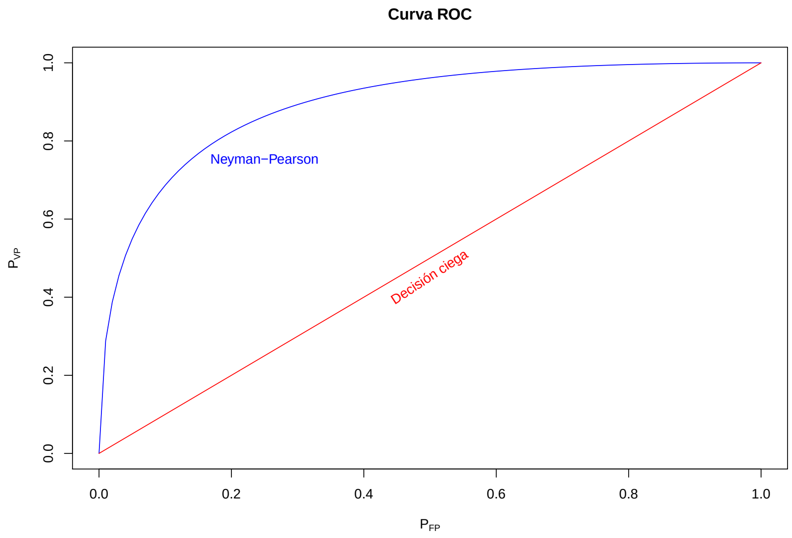
$$P_{FP} = P(\text{ rechazar } H0|H0 \text{ cierta}) = \alpha,$$
$$P_{VP} = P(\text{ rechazar } H0|H0 \text{ cierta}) = \alpha.$$

  Es decir el punto de la curva ROC de la decisión ciega sería (α, α).

- Por tanto, la curva ROC correspondiente a la decisión ciega sería la diagonal del cuadrado $[0,1] \times [0,1]$.

- Entonces, cuanto más lejos esté la curva ROC de nuestra regla de decisión de la diagonal del cuadrado $[0,1] \times [0,1]$, más lejos estará de la decisión ciega y más óptima será.

Decisión ciega



La curva *Precision-Recall*

- Una alternativa a la curva ROC para medir el rendimiento de la regla de decisión es la curva Precision-Recall o curva PR. Se define de la forma siguiente:

Definición

Se define la precisión de una regla de decisión como:

$$\text{precisión} = \frac{P_{VP}}{P_{VP} + P_{FP}},$$

y el recall como:

$$\text{recall} = \frac{P_{VP}}{P_{VP} + P_{FN}} = P_{VP}.$$

La curva *Precision-Recall*

- Es decir, la precisión sería el número de verdaderos positivos entre todos los verdaderos positivos y falsos positivos o el total de positivos detectados.

La curva *Precision-Recall*

- Es decir, la precisión sería el número de verdaderos positivos entre todos los verdaderos positivos y falsos positivos o el total de positivos detectados.
- Una precisión alta significa que entre todos los positivos detectados, la mayoría son verdaderos positivos. Es decir, lo que la regla de decisión afirma que son positivos, en realidad lo son o la regla es muy fiable. En este caso, si rechazamos la hipotesis nula, probablemente tomemos una buena decisión.

La curva *Precision-Recall*

- Es decir, la precisión sería el número de verdaderos positivos entre todos los verdaderos positivos y falsos positivos o el total de positivos detectados.

- Una precisión alta significa que entre todos los positivos detectados, la mayoría son verdaderos positivos. Es decir, lo que la regla de decisión afirma que son positivos, en realidad lo son o la regla es muy fiable. En este caso, si rechazamos la hipotesis nula, probablemente tomemos una buena decisión.

- Una precisión baja significa que entre todos los positivos detectados, la mayoría en realidad no lo son. Una precisión baja puede ocurrir cuando uno está demasiado ansioso para rechazar H_1. En este caso, si rechazamos la hipótesis nula, probablemente no sea una buena decisión.

La curva *Precision-Recall*

- El recall sería simplemente la tasa de verdaderos positivos.

La curva *Precision-Recall*

- El recall sería simplemente la tasa de verdaderos positivos.
- La diferencia fundamental entre la precisión y el recall es el denominador. En el caso, de la precisión, dividimos entre todos los positivos detectados; en cambio, en el caso del recall, dividimos entre todos los positivos verdaderos.

La curva *Precision-Recall*

- El recall sería simplemente la tasa de verdaderos positivos.
- La diferencia fundamental entre la precisión y el recall es el denominador. En el caso, de la precisión, dividimos entre todos los positivos detectados; en cambio, en el caso del recall, dividimos entre todos los positivos verdaderos.
- Un recall alto significa que la regla de decisión es muy fiable en detectar positivos o rechazar la hipótesis nula H_0. Esto ocurre cuando elegimos un valor α muy bajo. Sin embargo, en este caso, tendríamos una precisión baja.

La curva *Precision-Recall*

- El recall sería simplemente la tasa de verdaderos positivos.
- La diferencia fundamental entre la precisión y el recall es el denominador. En el caso, de la precisión, dividimos entre todos los positivos detectados; en cambio, en el caso del recall, dividimos entre todos los positivos verdaderos.
- Un recall alto significa que la regla de decisión es muy fiable en detectar positivos o rechazar la hipótesis nula H_0. Esto ocurre cuando elegimos un valor α muy bajo. Sin embargo, en este caso, tendríamos una precisión baja.
- Un recall bajo significa que aceptamos la hipótesis nula demasiadas veces y por tanto, casi nunca detectamos un positivo. Sin embargo, tener un recall bajo significa tener una precisión alta ya que casi nunca rechazamos la hipótesis nula a no ser que tengamos evidencia clara.

La curva *Precision-Recall*

- La curva PR consiste en realizar un gráfico de la precisión (eje Y) en función del recall o tasa de verdaderos positivos (eje X).