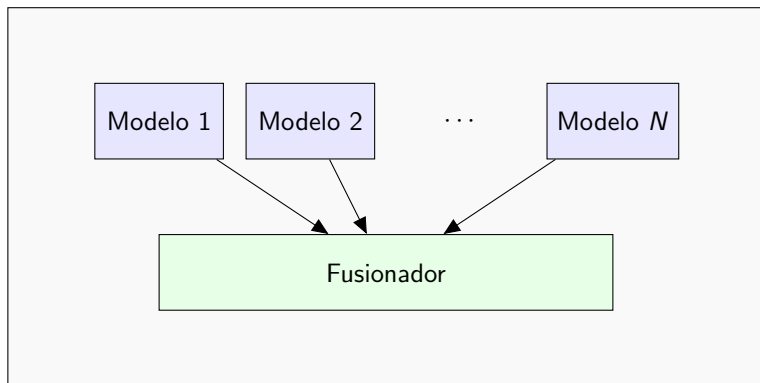


# Método de evaluación para modelo integrante de un sistema clasificador (Análisis de peor caso)

# Introducción

Se tiene un sistema clasificador con distintos modelos en paralelo que pueden ver features distintas de los datos. Luego se fusionan sus resultados mediante algún método, resultando en un único resultado del sistema. Está enfocado particularmente en clasificación binaria.

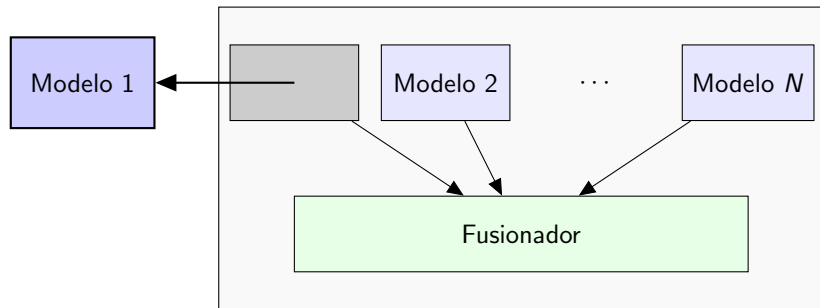
## Sistema Clasificador



# Introducción

Se busca evaluar un modelo para ver como este impacta en el rendimiento del sistema. En otras palabras, se quiere poder evaluar que tan bien se espera que funcione el sistema dado que cierto modelo sea usado.

# Introducción



# Información disponible

Datos de los que se dispone para el problema:

- ▶ El sistema incompleto, sin el modelo
- ▶ Datos para la evaluación del sistema
- ▶ Una función que asocia las confusiones posibles del sistema a costos
- ▶ El modelo que se pretende evaluar
- ▶ Datos para la evaluación del modelo

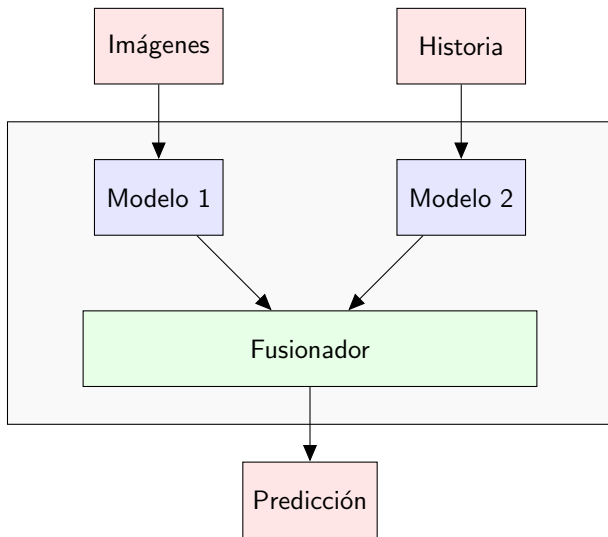
# Problema

Para evaluar el modelo, una opción es correr el sistema con el modelo y ver que resultados se obtienen. Para esto es necesario que quien tenga el modelo, también tenga acceso al sistema. Incluso aunque eso fuera posible, se podría querer evitar la ejecución del sistema completo para cada modelo posible que se busque evaluar. El objetivo es proponer un método que sea de utilidad para comparar potenciales modelos sin la necesidad de evaluar el sistema nuevamente para cada uno.

# Sistema ejemplo

Se usará como ejemplo el siguiente sistema. Se tienen datos de tipo imágenes médicas y historia clínica. El sistema consiste de dos modelos distintos, cada uno con su veredicto que son luego fusionados con algún método. El modelo a evaluar en este ejemplo es el modelo que recibe historias clínicas.

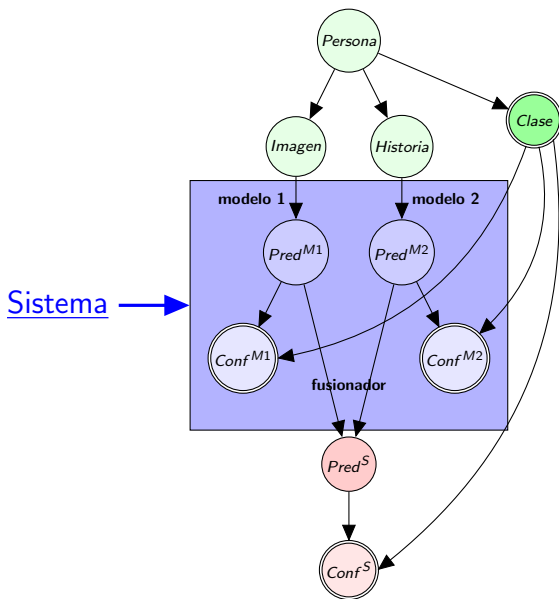
# Sistema ejemplo



# Redes Bayesianas

La forma que se va a utilizar para modelar el problema es mediante una red bayesiana causal. Una red bayesiana causal es un DAG que se usa para representar distribuciones probabilísticas sobre variables no independientes. Se tienen nodos para representar entidades, y aristas que representan causalidad entre esas entidades. Los nodos con doble círculo son determinísticos dados los padres.

# Redes Bayesianas



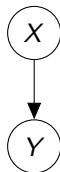
# Intervención

En las redes bayesianas existe una operación interesante llamada intervención. Esta operación le asigna a la variable de cierto nodo un valor particular, con probabilidad 1 en todos los casos. Esta operación se representa en una consulta probabilística con el operador **do**.

# Intervención

Un ejemplo del operador **do**:

- ▶  $P(Y = y | X = x)$
- ▶  $P(Y = y | do(X = x))$



La consulta sin **do** es un condicional normal, donde miramos la probabilidad de  $Y = y$  solo para los casos en los que  $X = x$ .

La consulta con el operador **do** se realiza sobre toda la distribución, asumiendo que  $X$  vale  $x$  siempre.

# Estimación del costo del sistema

Para poder comparar entre modelos se separa el proceso en dos partes:

Primero es evaluado el sistema sin el modelo. Se calculan las probabilidades de error y acierto si el modelo hubiera fallado o acertado siempre, usando la operación de intervención.

Luego al tener esos valores, se ejecuta un modelo en particular y se intenta estimar alguna métrica del sistema asociada a ese modelo.

# Estimación del costo del sistema

- ▶ Idealmente, se buscaría estimar el resultado de la evaluación del sistema si se hubiera corrido con el modelo. Esto no es posible ya que dados distintos modelos con los mismos resultados en su matriz de confusión, es posible que interactúen de distintas formas con el sistema y llevar a resultados finales muy distintos.
- ▶ Es por esto que la métrica de interés va a ser que dada la matriz de confusión de un modelo candidato para un conjunto de instancias, estimar el peor resultado posible de la evaluación del sistema sobre esas instancias.

# Notación

Para definir el problema formalmente se introduce la notación  $m_{ij}^{f,D}$  para denotar un valor de una matriz de confusión:

$$m_{ij}^{f,D} = |\{d \in D \mid f(d_1) = i \wedge d_2 = j\}|$$

- ▶  $i$  es la clase real
- ▶  $j$  es la clase que predice el modelo  $f$  sobre el dato
- ▶  $D$  es el dataset utilizado, conteniendo un conjunto de pares  $d_1$  instancia,  $d_2$  clase real y su valor por defecto que representa todos los datos disponibles es  $D$
- ▶  $f$  es el modelo usado y puede valer:
  - ▶  $M$  como el modelo a analizar
  - ▶  $S$  como el sistema completo con el modelo
  - ▶  $S_{M=X}$  como el sistema con un modelo intervenido que predice siempre la clase  $X$  para cualquier instancia

# Especificación del problema

Se espera lo siguiente sobre el comportamiento del sistema:

$$\begin{aligned} \blacktriangleright \{d \in D \mid d_2 = 0 \wedge S(d_1) = 0 \wedge M(d_1) = 1\} \subseteq \\ \{d \in D \mid d_2 = 0 \wedge S(d_1) = 0 \wedge M(d_1) = 0\} \end{aligned}$$

\\ Agregar evidencia solo puede mejorar la predicción

# Especificación del problema

Se espera lo siguiente sobre el comportamiento del sistema:

$$\begin{aligned} \text{▶ } \{d \in D \mid d_2 = 0 \wedge S(d_1) = 1 \wedge M(d_1) = 0\} \subseteq \\ \{d \in D \mid d_2 = 0 \wedge S(d_1) = 1 \wedge M(d_1) = 1\} \end{aligned}$$

\\ Restar evidencia solo puede empeorar la predicción

# AssEspecificación del problemaassumptions

Se espera lo siguiente sobre el comportamiento del sistema:

- ▶  $\{d \in D \mid d_2 = 0 \wedge S(d_1) = 0 \wedge M(d_1) = 1\} \subseteq \{d \in D \mid d_2 = 0 \wedge S(d_1) = 0 \wedge M(d_1) = 0\}$
- ▶  $\{d \in D \mid d_2 = 0 \wedge S(d_1) = 1 \wedge M(d_1) = 0\} \subseteq \{d \in D \mid d_2 = 0 \wedge S(d_1) = 1 \wedge M(d_1) = 1\}$
- ▶  $\{d \in D \mid d_2 = 1 \wedge S(d_1) = 1 \wedge M(d_1) = 0\} \subseteq \{d \in D \mid d_2 = 1 \wedge S(d_1) = 1 \wedge M(d_1) = 1\}$
- ▶  $\{d \in D \mid d_2 = 1 \wedge S(d_1) = 0 \wedge M(d_1) = 1\} \subseteq \{d \in D \mid d_2 = 1 \wedge S(d_1) = 0 \wedge M(d_1) = 0\}$

\\ Lo mismo para Clase Negative

# Especificación del problema

Se espera lo siguiente sobre la función de costo asociada a las confusiones del sistema:

- ▶  $Costo_{00}^S \leq Costo_{01}^S$

- ▶  $Costo_{11}^S \leq Costo_{10}^S$

Los fallos tienen mayor costo que los aciertos.

# Especificación del problema

Datos disponibles:

	Pred Pos	Pred Neg
Clase Pos	$m_{0,0}^{S_{M=P}, D}$	$m_{0,1}^{S_{M=P}, D}$
Clase Neg	$m_{1,0}^{S_{M=P}, D}$	$m_{1,1}^{S_{M=P}, D}$

Table: Matriz de confusión del sistema si Modelo = Positive

	Pred Pos	Pred Neg
Clase Pos	$m_{0,0}^{S_{M=N}, D}$	$m_{0,1}^{S_{M=N}, D}$
Clase Neg	$m_{1,0}^{S_{M=N}, D}$	$m_{1,1}^{S_{M=N}, D}$

Table: Matriz de confusión del sistema si Modelo = Negative

# Especificación del problema

Otros datos disponibles:

	Positive	Negative
#Instancias	P	N

Table: Cantidad de instancias por clase

	Pred Pos	Pred Neg
Clase Pos	$m_{0,0}^{M,D}$	$m_{0,1}^{M,D}$
Clase Neg	$m_{1,0}^{M,D}$	$m_{1,1}^{M,D}$

Table: Matriz de confusión del modelo

# Especificación del problema

Dados esos datos, se busca generar la peor matriz de confusión posible del sistema que sea válida (peor caso).

	Prediccion Positive	Prediccion Negative
Clase Positive	$m_{0,0}^{S,D}$	$m_{0,1}^{S,D}$
Clase Negative	$m_{1,0}^{S,D}$	$m_{1,1}^{S,D}$

Table: Matriz de confusión del sistema

# Especificación del problema

La peor es la que maximiza:

$$\begin{aligned} \blacktriangleright & m_{0,0}^{S,D} * Costo_{00}^S + m_{0,1}^{S,D} * Costo_{01}^S + \\ & m_{1,0}^{S,D} * Costo_{10}^S + m_{1,1}^{S,D} * Costo_{11}^S \end{aligned}$$

Como los costos asociados a las confusiones de aciertos son menores a las de fallos, es equivalente a maximizar:

$$\blacktriangleright m_{0,1}^{S,D} + m_{1,0}^{S,D}$$

# Cotas independientes al modelo

Si se logra acotar de forma exacta las confusiones del sistema, particularmente hallar una cota superior para  $m_{0,1}^{S,D}$  y  $m_{1,0}^{S,D}$ , podemos tomar ese valor máximo como el valor de la confusión en peor caso.

Para las confusiones restantes  $m_{0,0}^{S,D}$  y  $m_{1,1}^{S,D}$  se toman las instancias restantes de la clase que deberían corresponder a su cota inferior.

# Cotas independientes al modelo

Cotas a las confusiones del sistema (independientes al modelo):

$$\blacktriangleright m_{0,0}^{S,D} + m_{0,1}^{S,D} = P$$

$$\blacktriangleright m_{1,1}^{S,D} + m_{1,0}^{S,D} = N$$

\\ Total de instancias por clase

# Cotas independientes al modelo

Cotas a las confusiones del sistema (independientes al modelo):

$$\blacktriangleright m_{0,0}^{S,D} + m_{0,1}^{S,D} = P$$

$$\blacktriangleright m_{1,1}^{S,D} + m_{1,0}^{S,D} = N$$

$$\blacktriangleright m_{0,0}^{S_{M=N},D} \leq m_{0,0}^{S,D} \leq m_{0,0}^{S_{M=P},D} \quad \backslash \backslash \quad \text{Más aciertos que cuando el modelo falla siempre, menos aciertos que cuando el modelo acierta siempre}$$

# Cotas independientes al modelo

Cotas a las confusiones del sistema (independientes al modelo):

$$\blacktriangleright m_{0,0}^{S,D} + m_{0,1}^{S,D} = P$$

$$\blacktriangleright m_{1,1}^{S,D} + m_{1,0}^{S,D} = N$$

$$\blacktriangleright m_{0,0}^{S_{M=N},D} \leq m_{0,0}^{S,D} \leq m_{0,0}^{S_{M=P},D}$$

$$\blacktriangleright m_{0,1}^{S_{M=P},D} \leq m_{0,1}^{S,D} \leq m_{0,1}^{S_{M=N},D}$$

\\ Más fallos que cuando el modelo acierta siempre, menos fallos que cuando el modelo falla siempre

# Cotas independientes al modelo

Cotas a las confusiones del sistema (independientes al modelo):

$$\blacktriangleright m_{0,0}^{S,D} + m_{0,1}^{S,D} = P$$

$$\blacktriangleright m_{1,1}^{S,D} + m_{1,0}^{S,D} = N$$

$$\blacktriangleright m_{0,0}^{S_{M=N},D} \leq m_{0,0}^{S,D} \leq m_{0,0}^{S_{M=P},D}$$

$$\blacktriangleright m_{0,1}^{S_{M=P},D} \leq m_{0,1}^{S,D} \leq m_{0,1}^{S_{M=N},D}$$

$$\blacktriangleright m_{1,0}^{S_{M=N},D} \leq m_{1,0}^{S,D} \leq m_{1,0}^{S_{M=P},D}$$

$$\blacktriangleright m_{1,1}^{S_{M=P},D} \leq m_{1,1}^{S,D} \leq m_{1,1}^{S_{M=N},D}$$

\\ Lo mismo para la clase Negative

# Cotas independientes al modelo

Hasta ahora se tienen unos valores que acotan superior e inferiormente a cada confusión del sistema. Este rango es independiente del modelo y potencialmente muy laxo.

A continuación se va a analizar de forma más cercana los posibles resultados de las instancias en base a lo que se sabe del modelo y del comportamiento del sistema para poder acotar de forma más ajustada.

## Análisis de peor caso

La siguiente tabla tiene para con conjunto de instancias de ejemplo de clase Positive, que resultado tendría el sistema en caso del modelo predecir Positive o Negative.

Se marca con una X las instancias donde el sistema falla. La suma de la cantidad de X por columna son los valores con intervención mencionados anteriormente en  $m_{0,1}^{S_{M=P},D}$  y  $m_{0,1}^{S_{M=N},D}$ . Se busca ilustrar que casos no son válidos en base a las assumptions planteadas.

El mismo análisis es válido para la clase Negative.

# Análisis de peor caso

Recordando la assumption 2 de la clase Positive:

$$\{d \in D \mid d_2 = 0 \wedge S(d_1) = 0 \wedge M(d_1) = 1\} \subseteq \\ \{d \in D \mid d_2 = 0 \wedge S(d_1) = 0 \wedge M(d_1) = 0\}$$

	$m_{01}^{S^{M=P},i}$	$m_{01}^{S^{M=N},i}$	Es Válido
Instancia 1			
Instancia 2	X		
Instancia 3		X	
Instancia 4	X	X	

**Table:** Resultados posibles en instancias de clase Positive

# Análisis de peor caso

Recordando la assumption 2 de la clase Positive:

$$\{d \in D \mid d_2 = 0 \wedge S(d_1) = 0 \wedge M(d_1) = 1\} \subseteq \\ \{d \in D \mid d_2 = 0 \wedge S(d_1) = 0 \wedge M(d_1) = 0\}$$

	$m_{01}^{S^{M=P},i}$	$m_{01}^{S^{M=N},i}$	Es Válido
Instancia 1			SI
Instancia 2	X		
Instancia 3		X	
Instancia 4	X	X	

**Table:** Resultados posibles en instancias de clase Positive

# Análisis de peor caso

Recordando la assumption 2 de la clase Positive:

$$\{d \in D \mid d_2 = 0 \wedge S(d_1) = 0 \wedge M(d_1) = 1\} \subseteq \\ \{d \in D \mid d_2 = 0 \wedge S(d_1) = 0 \wedge M(d_1) = 0\}$$

	$m_{01}^{S^{M=P},i}$	$m_{01}^{S^{M=N},i}$	Es Válido
Instancia 1			SI
Instancia 2	X		NO
Instancia 3		X	
Instancia 4	X	X	

**Table:** Resultados posibles en instancias de clase Positive

# Análisis de peor caso

Recordando la assumption 2 de la clase Positive:

$$\{d \in D \mid d_2 = 0 \wedge S(d_1) = 0 \wedge M(d_1) = 1\} \subseteq \\ \{d \in D \mid d_2 = 0 \wedge S(d_1) = 0 \wedge M(d_1) = 0\}$$

	$m_{01}^{S^{M=P},i}$	$m_{01}^{S^{M=N},i}$	Es Válido
Instancia 1			SI
Instancia 2	X		NO
Instancia 3		X	SI
Instancia 4	X	X	

**Table:** Resultados posibles en instancias de clase Positive

# Análisis de peor caso

Recordando la assumption 2 de la clase Positive:

$$\{d \in D \mid d_2 = 0 \wedge S(d_1) = 0 \wedge M(d_1) = 1\} \subseteq \\ \{d \in D \mid d_2 = 0 \wedge S(d_1) = 0 \wedge M(d_1) = 0\}$$

	$m_{01}^{S^{M=P},i}$	$m_{01}^{S^{M=N},i}$	Es Válido
Instancia 1			SI
Instancia 2	X		NO
Instancia 3		X	SI
Instancia 4	X	X	SI

**Table:** Resultados posibles en instancias de clase Positive

## Análisis de peor caso

Ahora sabiendo de que formas se pueden comportar las instancias, se observarán distintas posibles combinaciones de confusiones del sistema para las instancias de clase Positive, fijando los resultados del modelo. Si una fila es verde el modelo predice Positive y si es roja predice Negative. Al fijar el modelo se fijan la cantidad de predicciones Positive y Negative, pero hay muchas formas de adquirir la misma confusión que llevan a distintas confusiones del sistema.

La cantidad total de X en la columna de Positive es  $m_{0,1}^{S_{M=P},D}$  y la cantidad total de X en la columna de Negative es  $m_{0,1}^{S_{M=N},D}$ .

# Análisis de peor caso

	$m_{01}^{S^{M=P},i}$	$m_{01}^{S^{M=N},i}$	$m_{01}^{S,i}$
Instancia 1		X	
Instancia 2		X	
Instancia 3			
Instancia 4	X	X	X
Instancia 5			
Instancia 6		X	X
Instancia 7		X	X
Instancia 8		X	X

Table: Ejemplo Clase Positive 1

Cantidad de True Positives totales: 4  
Cantidad de False Negatives totales: 4

# Análisis de peor caso

	$m_{01}^{S^{M=P},i}$	$m_{01}^{S^{M=N},i}$	$m_{01}^{S,i}$
Instancia 1		X	
Instancia 2		X	
Instancia 3			
Instancia 4	X	X	X
Instancia 5		X	X
Instancia 6			
Instancia 7		X	
Instancia 8		X	X

Table: Ejemplo Clase Positive 2

Cantidad de True Positives totales: 5  
Cantidad de False Negatives totales: 3

# Análisis de peor caso

	$m_{01}^{S^{M=P},i}$	$m_{01}^{S^{M=N},i}$	$m_{01}^{S,i}$
Instancia 1		X	X
Instancia 2		X	
Instancia 3			
Instancia 4	X	X	X
Instancia 5			
Instancia 6		X	X
Instancia 7		X	X
Instancia 8		X	X

Table: Ejemplo Clase Positive 3

Cantidad de True Positives totales: 3

Cantidad de False Negatives totales: 5

# Análisis de peor caso

Se puede observar como para 3 modelos ejemplo con la misma matriz de confusión, se obtienen 3 resultados de sistema distintos. La causa de esto son las instancias con una sola X, sobre las cuales si el modelo predice Positive, el resultado del sistema es un True Positive y si el modelo predice Negative, el resultado del sistema es un False Negative.

# Análisis de peor caso

Ahora se busca saber cual es la cantidad de False Negatives máxima y la cantidad de True Positive máxima luego de fijar el modelo, y así ajustar la cota anterior.

Particularmente se va a analizar solo la cota superior para False Negative que es la de interés, al brindar el peor caso que se busca.

# Análisis de peor caso

Las confusiones del sistema se pueden particionar de la siguiente manera:

$$\blacktriangleright m_{0,0}^{S,D} = m_{0,0}^{S,D_{M=P}} + m_{0,0}^{S,D_{M=N}}$$

$$\blacktriangleright m_{0,1}^{S,D} = m_{0,1}^{S,D_{M=P}} + m_{0,1}^{S,D_{M=N}}$$

Con el objetivo de separar en casos por resultado del modelo y acotar ambos casos por separado.

# Análisis de peor caso

La cantidad máxima de False Negatives ocurre cuando:

- ▶  $m_{0,0}^{S,D_{M=P}}$  tiene la máxima cantidad de instancias con doble X posibles.
- ▶  $m_{0,0}^{S,D_{M=N}}$  tiene la máxima cantidad de instancias con una X posibles.

De esta forma todas las instancias con una X que sea posible considerar como False Negatives del modelo van a serlo.

# Análisis de peor caso

Escribiendo formalmente la cota superior de fallos para la clase Positive:

$$\blacktriangleright m_{0,1}^{S,D_{M=P}} = \text{Min}(m_{0,0}^{M,D}, m_{0,1}^{S_{M=P},D})$$

$$\blacktriangleright m_{0,1}^{S,D_{M=N}} = \text{Min}(m_{0,1}^{M,D}, m_{0,1}^{S_{M=N},D} - \text{Min}(m_{0,0}^{M,D}, m_{0,1}^{S_{M=P},D}))$$

# Análisis de peor caso

Definiendo la cota superior a los fallos del sistema, e inferior a los aciertos, dado que se saben las confusiones del modelo:

$$m_{0,0}^{S,D} \geq (m_{0,0}^{M,D} - \text{Min}(m_{0,0}^{M,D}, m_{0,1}^{S_{M=P},D})) + \\ (m_{0,1}^{M,D} - \text{Min}(m_{0,1}^{M,D}, m_{0,1}^{S_{M=N},D} - \text{Min}(m_{0,0}^{M,D}, m_{0,1}^{S_{M=P},D})))$$

$$m_{0,1}^{S,D} \leq \text{Min}(m_{0,0}^{M,D}, m_{0,1}^{S_{M=P},D}) + \\ \text{Min}(m_{0,1}^{M,D}, m_{0,1}^{S_{M=N},D} - \text{Min}(m_{0,0}^{M,D}, m_{0,1}^{S_{M=P},D}))$$

# Análisis de peor caso

A su vez, el mismo análisis es válido para la clase Negative, llegando a las siguientes cotas:

Cota superior a los fallos del sistema, e inferior a los aciertos, dado que se saben las confusiones del modelo:

$$\begin{aligned} m_{1,0}^{S,D} &\leq \text{Min}(m_{1,1}^{M,D}, m_{1,0}^{S_{M=N},D}) + \\ &\quad \text{Min}(m_{1,0}^{M,D}, m_{1,0}^{S_{M=P},D} - \text{Min}(m_{1,1}^{M,D}, m_{1,0}^{S_{M=N},D})) \\ m_{1,1}^{S,D} &\geq (m_{1,1}^{M,D} - \text{Min}(m_{1,1}^{M,D}, m_{1,0}^{S_{M=N},D})) + \\ &\quad (m_{1,0}^{M,D} - \text{Min}(m_{1,0}^{M,D}, m_{1,0}^{S_{M=P},D} - \text{Min}(m_{1,1}^{M,D}, m_{1,0}^{S_{M=N},D}))) \end{aligned}$$

## Análisis de peor caso

Como el objetivo es encontrar el máximo para  $m_{0,1}^{S,D}$  y  $m_{1,0}^{S,D}$  se toma el valor de esas confusiones igual a sus cotas superiores, y las confusiones  $m_{0,0}^{S,D}$  y  $m_{1,1}^{S,D}$  asociadas iguales a sus cotas inferiores.

# Análisis de peor caso

Por último, se multiplican cada confusión por su costo asociado para calcular el costo del sistema en su peor caso:

$$EvalPeorCaso^S = m_{0,0}^{S,D} * Costo_{00}^S + m_{0,1}^{S,D} * Costo_{01}^S + \\ m_{1,0}^{S,D} * Costo_{10}^S + m_{1,1}^{S,D} * Costo_{11}^S$$

# Datasets distintos para evaluación

Hasta ahora se trabajó asumiendo que los datos en los que se evalúan las intervenciones sobre el sistema, y los datos con los que se evalúa el modelo son los mismos, y se logró encontrar una función de costo de peor caso para el sistema. Originalmente se buscaba conseguir una función de costo que fuera útil también para casos donde los datos en los que se evalúa el sistema intervenido y los datos con los que se evalúa el modelo fueran distintos.

# Datasets distintos para evaluación

- ▶ La función de costo de peor caso solo utiliza el accuracy de los resultados del sistema con intervenido. Mientras ese accuracy medido sea el mismo que el accuracy que tendría el sistema (intervenido) en los datos con los que se evalúa el modelo, la función sigue calculando el peor caso correctamente.
- ▶ Por ley de los grandes números, si la predicción del sistema en una instancia cualquiera es una variable aleatoria iid con media en el accuracy, mientras la cantidad de datos de cada conjunto de evaluación sea lo suficientemente grande, el accuracy de cada conjunto de evaluación va a tender a la media poblacional (accuracy real) y por lo tanto la medición del primer conjunto de datos con intervención va a ser la misma que el accuracy efectivo para el segundo conjunto de datos.

# Experimentación

Se evaluó la función de costo para el modelo ejemplo anterior, donde el sistema recibe como entrada imágenes para el modelo 1 y historia clínica para el modelo 2. Luego un fusionador recibe los resultados de clasificación de ambos modelos y en base a alguna operación lógica determina el resultado final del sistema (OR o AND).

Este sistema fue simulado con variables aleatorias Bernoulli para cada instancia y cada clase. Se tiene una variable por par de modelo-clase y con correlación entre variables representantes de distintos modelos para la misma clase. Esto es para simular como los modelos pueden tener resultados no independientes dada una instancia.

# Experimentación

La experimentación fue separada en dos etapas.

La primera etapa compuesta por una evaluación inicial del sistema sin el modelo de historias clínicas. Aquí son estimadas las precisiones del sistema al intervenir en los resultados del modelo para cada instancia:

▶  $m_{0,0}^{S_{M=P},D}$

▶  $m_{0,1}^{S_{M=P},D}$

▶  $m_{1,0}^{S_{M=P},D}$

▶  $m_{1,1}^{S_{M=P},D}$

▶  $m_{0,0}^{S_{M=N},D}$

▶  $m_{0,1}^{S_{M=N},D}$

▶  $m_{1,0}^{S_{M=N},D}$

▶  $m_{1,1}^{S_{M=N},D}$

# Experimentación

La segunda etapa corre el modelo en un nuevo conjunto de datos distinto, generado con la misma distribución que el anterior y calcula las confusiones de cada resultado. Luego evalúa la función de costo de peor caso para ese conjunto de datos.

Por otro lado, simula lo que hubiera pasado realmente con el sistema si esos resultados del modelo fuesen usados.

# Experimentación

Al correr el experimento se pudo ver que la función de peor caso, al ser usada en sistemas simulados con una muy mala interacción con el modelo en cuestión, puede llegar a subestimar el costo del sistema. Esto se debe a que cuando el sistema es evaluado sobre el nuevo conjunto de datos, este puede funcionar peor que lo anteriormente medido en la etapa 1 de los experimentos y que luego los cálculos no sean del todo precisos.

# Experimentación

Al ver esto se corrieron dos experimentos distintos:

Uno simulando un modelo con valores fijos de precision en ambos modelos y con una relación entre los resultados del modelo y del sistema bastante desfavorable. Su objetivo es analizar la probabilidad de que la estimación del peor caso de la función de costo sea realmente peor que la ejecución del sistema real y como esta probabilidad se comporta al cambiar la cantidad de instancias de testing de ambas fases del experimento.

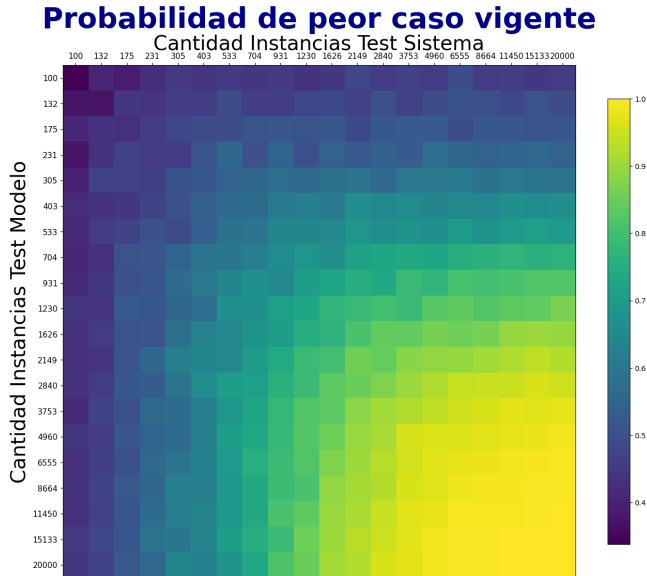
$$\blacktriangleright P(EvalPeorCaso^S > m_{0,0}^{S,D} * Costo_{00}^S + m_{0,1}^{S,D} * Costo_{01}^S + m_{1,0}^{S,D} * Costo_{10}^S + m_{1,1}^{S,D} * Costo_{11}^S)$$

# Experimentación

El otro simulando un sistema con precisiones de ambos modelos y con una interacción entre el sistema y el modelo aleatorias, cuyo objetivo es evaluar que tanto se llega a acercar la evaluación real del sistema a la estimación de peor caso.

$$\blacktriangleright \frac{Costo_{00}^S * m_{0,0}^{S,D} + Costo_{01}^S * m_{0,1}^{S,D} + Costo_{11}^S * m_{1,1}^{S,D} + Costo_{10}^S * m_{1,0}^{S,D}}{EvalPeorCaso^S}$$

# Resultados

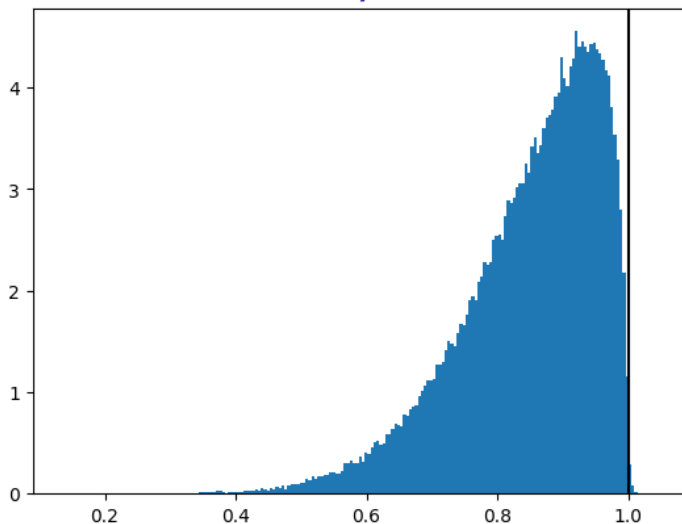


# Resultados

En el gráfico anterior se ven los resultados del primer experimento, donde se grafica la probabilidad de que el peor caso sea efectivamente mayor a la evaluación del sistema. El sistema es uno compuesto por ambos modelos con accuracies de 50% para ambas clases, con correlación de 0.95 para instancias de clase Positive y -0.95 de clase Negative y un fusionador de tipo OR. Se puede apreciar como la probabilidad se acerca rápidamente a 1 a medida que ambas cantidades de instancias aumentan. Se ejecutó 1000 veces para cada par de cantidades de instancias.

# Resultados

## Evaluacion/Peor Caso



# Resultados

En el gráfico anterior se ven los resultados del segundo experimento, corrido con una cantidad de instancias fija en 20000 para cada experimento. Se corrió el experimento 100000 veces para graficar la distribución de la relación entre ambos valores. Aquí se puede apreciar como la relación entre el costo del sistema y la estimación de peor caso tiene como cota superior el valor 1, salvo algunos valores atípicos que levemente lo superan.

# Preguntas?