

# Practica 6. Naive Bayes

---

DRA. CONSUELO VARINIA GARCÍA MENDOZA

# Especificaciones

---

- En equipo de máximo 3 integrantes
- Los dataset utilizados para esta práctica serán *iris.csv* y *emails.csv*
  - En el dataset *emails.csv*
    - La primera columna indica el id del correo
    - La última columna indica si el correo es spam o no
    - El resto de las columnas son las palabras más comunes en todos los correos
  - En el dataset *iris.csv*
    - Las primeras 4 columnas son las características de las instancias
    - La última columna es la clase
- Utilizando ambos dataset realiza lo siguiente:
  - Carga el dataset
  - Crea un conjunto de entrenamiento del 70% de las instancias y el conjunto de pruebas con el 30%(set *random\_state = 0*)
- Utilizando el conjunto de entrenamiento crea un conjunto de validación con 3 pliegues

# Especificaciones

---

- Utilizando el conjunto de validación realiza lo siguiente con cada pliegue:
  - Utiliza Naive Bayes con distribución normal para entrenar y prueba el modelo en cada pliegue
    - Determina la exactitud de cada pliegue
    - Determina el promedio de exactitud de los 3 pliegues
  - Utiliza Multinomial con distribución normal para entrenar y prueba el modelo en cada pliegue
    - Determina la exactitud de cada pliegue
    - Determina el promedio de exactitud de los 3 pliegues

# Evidencias

---

- Código fuente
- Un PDF con la siguiente información:
  - Tabla con los resultados

Dataset	Clasificador	Exactitud promedio de los 3 pliegues	Exactitud obtenida con el conjunto de prueba (30%)
emails.csv	Gaussiano		
emails.csv	Multinomial		
iris.csv	Gaussiano		
iris.csv	Multinomial		

# Evidencias

---

- Reporte de clasificación y matriz de confusión del conjunto de entrenamiento(70%) de los dataset (iris y email) con Gauss y Mulinomial (4 reportes de entrenamiento)
- Reporte de clasificación y matriz de confusión del conjunto de prueba(30%) de los dataset (iris y email) con Gauss y Mulinomial (4 reportes de prueba)
- Conclusiones generales de la práctica mencionando las dificultades encontradas y propuestas para mejorar los resultados obtenidos

		PREDICTION	
		true	false
REALITY	true	a	b
	false	c	d

$$F_1 = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}}$$

<b>accuracy:</b> “what percent of the prediction were correct?”	$\frac{(a+d)}{(a+b+c+d)}$
<b>precision:</b> “what percent of positive predictions were correct?”	$\frac{(a)}{(a+c)}$
<b>recall:</b> “what percent of positive cases were caught?”	$\frac{(a)}{(a+b)}$