Este cuestionario se contesta a partir de la notebook “NBM y NBG para Fashion MNIST”

**Dataset Fashion MNIST**

1. Considerando el dataset de Fashion MNIST, si se modela cada pixel como una variable aleatoria continua, ¿cuántas variables aleatorias se samplean para generar una imagen?

Respuesta: 784

1. ¿Las clases del dataset están balanceadas? Respuesta:SI
2. Dividir los datos en train, validation y test, en un ratio 5:1:1, usando **random\_state=42,** de forma estratificada por clases. ¿Cuántas imágenes de cada clase hay en el set de train? Respuesta: 5000

**Naive Bayes Multinomial**

1. El cálculo de likelihoods y probabilidades a priori, es decir, el cálculo de los parámetros del modelo de NBM, ¿con qué subset del dataset se realiza?

Respuesta: train set

1. Construir un histograma de 255 bins para la distribución del nivel de gris del pixel [15,15] en imágenes de clase “Dress”. ¿Cuál es la frecuencia del valor de gris xg=147?

Respuesta:15

1. ¿Cuál es la probailidad a priori de la clase "Dress"? Respuesta\_ 0.1
2. ¿Cuál es  el likelihood del nivel de gris 𝑥𝑔=147xg=147 para la clase 'Dress', utilizando 255 niveles de gris? Respuesta: 0.003
3. ¿Cuál es el likelihood del nivel de gris $x\_g=147$ para la clase 'Dress', utilizando 10 niveles de gris? Respuesta: 0.0306
4. ¿Cuál es la probabilidad a posteriori no normalizada de que un pixel con 𝑥𝑔=124xg=124 en la posición [17,17]pertenezca a la clase "Bag"? Redondeado a 5 decimales. Respuesta: 0.00024
5. Después de realizar la búsqueda de hiperparámetros descripta en la notebook, ¿cuál combinación presenta mejor accuracy en el set de validación?
6. Reentrenando con dicha combinación sobre los datos de train y de validación, ¿qué accuracy se obtiene en los datos de test?

**Naive Bayes Gaussiano**

1. ¿Cuál es la **media** del pixel (14,14) en la clase 'Bag' estimados a partir del set de train? Responder redondeado con tres decimales

Media: 159.329

1. ¿Y cuál el desvío estándar? Redondear con tres decimales

Desvío estandar: 63.316

1. ¿Cuál es el likelihood de observar en las imágenes de 'Trouser' un valor de gris 145 en el pixel (10,10)? Estimado con el set de train. Redondear a 5 decimales.

0.00547

1. ¿Cuál es la probabilidad a posteriori no normalizada de que un valor de 145 observado en el pixel (10,10) pertenezca a una imagen de Trouser? Estimado con el set de train. Redondear a 6 decimales.

0.000547

1. ¿Qué clase es más probable que haya generado una imagen con valor de gris 145 en el pixel (10,10)?

Coat

1. ¿Cómo afecta el hecho de que las clases están balanceadas a la hora de aplicar el criterio de máximum likelihood?

Texto libre: como todas las clases tienen la misma probabilidad a priori, se pueden comparar directamente los likelihoods.

1. ¿Cuántos parámetros tendrá el modelo de Naive Bayes Gaussiano dado que se asumió independencia entre los píxeles?

1568

(784 medias y 784 varianzas = 1568 parámetros)

1. Si no se asume independencia, es decir que las covarianzas entre variables no se desprecian: ¿cuántos parámetros tendría el modelo?

308504

Por un lado, las 784 medias. Por otro lado, de la matriz de covarianza sacamos (784\*785)/2 parámetros, porque es simétrica (es decir, los parámetros son las 784 varianzas de la diagonal y los 307720 valores de media matriz).

1. En el NB Gaussiano para distintos hiperparámetros, ¿Cuál es el mejor accuracy que obtuvo en validación? Redondear a tres decimales

0.672

1. ¿Con qué valor de hiperparámetros lo obtuvo?

0.1

1. ¿Qué accuracy obtiene en test? Redondear a cuatro decimales

0.6708

1. Si ajustásemos un modelo de NB gaussiano asumiendo que un grupo de 4 píxeles vecinos pertenece a la misma distribución gaussiana, ¿cuántos parámetros tendría el modelo?

392

1. El modelo de clasificación de Naive Bayes gaussiano, ¿es un modelo generativo o discriminativo?

Generativo