



[Course](#) > [Clasific...](#) > [Labora...](#) > Lab-01-1

Lab-01-1

Abrir notebook

Pueden bajar o clonar el repositorio y correrlo local:

https://github.com/lab-ml-itba/Laboratorio-01/blob/master/1-Lab01-Naive_Bayes.ipynb

Abrir notebook o bajar repositorio.

En este laboratorio se simularán QDA y Naive Bayes. Abra la notebook, ejecute las celdas y analice el código. Haga las distintas pruebas indicadas en la notebook.

Tiene dos datasets y dos modelos para hacer pruebas

Responda las preguntas a continuación:

Pregunta 1

0.0/1.0 point (graded)

Grafique los dos datasets e indique cuales de las preguntas son verdaderas

☐ Uno de los datasets es linealmente separable

☐ Ninguno de los datasets es linealmente separable

- ☐ Es de esperar que QDA tenga una mejor performance en `generate_elipse_data()` que Naive Bayes
- ☐ Es de esperar que Naive Bayes tenga una mejor performance en `generate_elipse_data()`
- ☐ `generate_elipse_data()` genera datos uniformemente distribuidos y asigna clase azul a aquellas observaciones que quedan dentro de una elipse.
- ☐ `generate_gaussians_distributions()` genera 2 distribuciones Gaussianas bi-variables con distinto vector de medias e igual matriz de covarianza

You have used 0 of 2 attempts

Las siguientes preguntas refieren al dataset generado con `generate_gaussians_distributions()`

Pregunta 2

0.0/1.0 point (graded)

Cuando da el accuracy (score) para Naive Bayes? (Escribirlo con tres posiciones decimales)

You have used 0 of 5 attempts

Pregunta 3

0.0/1.0 point (graded)

Cuando da el accuracy (score) para QDA? (Escribirlo con tres posiciones decimales)

You have used 0 of 5 attempts

Pregunta 4

0.0/1.0 point (graded)

Indicar cuales de las siguientes preguntas son verdaderas

- ☐ La mejora de performance de QDA sobre Naive Bayes se debe a que el dataset fue generado con una de las Gaussianas no independiente.
- ☐ Un modelo generativo es un buen modelo para este dataset
- ☐ La razón por la que queda una franja en el caso de QDA es por el hecho que la distribución de los rojos cae mas lento que la de los azules
- ☐ La razón por la que queda una franja en el caso de QDA es por el hecho que la distribución de los azules cae mas lento que la de los rojos
- ☐ Es más probable que un punto azul caiga en (1.25, 0.88) que uno rojo (para QDA)

You have used 0 of 2 attempts

Pregunta 5

0.0/1.0 point (graded)

Aplicar var_smoothing=1 para Naive Bayes

Cuales de las siguientes afirmaciones son correctas

☐ Las varianzas de las Gaussianas estimadas aumentan

☐ Las varianzas de las Gaussianas estimadas disminuyen

☐ El modelo es mas "seguro" al momento de clasificar

☐ El modelo es mas "inseguro" al momento de clasificar

Submit

You have used 0 of 2 attempts

Pregunta 6

0.0/1.0 point (graded)

Supongamos que a pesar de que la cantidad de puntos que tenemos para las dos clases es la misma, se sabe que la probabilidad de seleccionar la Gaussiana que genera azules es 0.3 y la de generar los rojos es 0.7. Cambiar las priors y entrenar nuevamente Naive Bayes

Cuales de las siguientes afirmaciones son correctas

☐ La estimación de la Gaussiana azul tiene menor altura que en el caso anterior

☐ La estimación de la Gaussiana roja tiene menor altura que en el caso anterior

☐ El umbral de decisión se corre hacia el lado de los rojos clasificando algunos rojos como azules

☐ El umbral de decisión se corre hacia el lado de los azules clasificando algunos azules como rojos

Submit

You have used 0 of 2 attempts

Las siguientes preguntas refieren al dataset generado con `generate_elipse_data()`

Pregunta 7

0.0/1.0 point (graded)

Cuando da el accuracy (score) para Naive Bayes? (Escribirlo con tres posiciones decimales) | `priors=None, var_smoothing=1e-09`

Submit

You have used 0 of 5 attempts

Pregunta 8

0.0/1.0 point (graded)

Cuando da el accuracy (score) para QDA? (Escribirlo con tres posiciones decimales)

Submit

You have used 0 of 5 attempts

Pregunta 9

0.0/1.0 point (graded)

Cambie el `var_smoothing=0.2` en Naive Bayes.

Cuales de las siguientes afirmaciones son correctas

- ☐ La incidencia del `var_smoothing` se ve más en la distribución estimada de las azules

- ☐ La incidencia del `var_smoothing` se ve más en la distribución estimada de las rojas
- ☐ El modelo nuevo clasifica más puntos como rojos cuando deberían ser azules
- ☐ El modelo nuevo clasifica más puntos como azules cuando deberían ser rojos

You have used 0 of 2 attempts

Pregunta 10

0.0/1.0 point (graded)

Verificar que en este caso, el dataset no esta balanceado. Hay mas de una clase que de otra. Mirar en `model.class_prior_` para ver cual es la proporción de rojos vs azules. Tambien puede calcularlas contando la cantidad de cada clase y dividiendo para el total. Si lo hace, verifique que coincide con las estimadas en `model.class_prior_`

Forzamos ahora las priors a `[0.5, 0.5]`

Cuales de las siguientes preguntas son verdaderas

- ☐ El nuevo modelo clasifica más puntos azules que el modelo anterior
- ☐ El nuevo modelo clasifica más puntos rojos que el modelo anterior
- ☐ El umbral de decisión no puede dar una elipse inclinada debido a Naive Bayes que asume independencia

You have used 0 of 2 attempts

Pregunta 11

0.0/1.0 point (graded)

Como sabemos, QDA da como solución un umbral cuadrática en el plano y supone que cada una de las clases fueron generados con la distribución Gaussiana.

Si tomamos las estimaciones Gaussianas obtenidas (para cada clase) luego de entrenar QDA y generamos igual cantidad de observaciones para cada clase que en el dataset original.

Indicar cuales de las siguientes afirmaciones son correctas

- ☐ Las distribuciones estimadas de los likelihoods son buenos modelos generadores de las observaciones.
- ☐ Tendremos muchos mas rojos en la zona central comparando con el dataset original
- ☐ Podríamos tener puntos azules fuera de la elipse con que fueron recortados los datos azules en el dataset original
- ☐ Los puntos van a quedar distribuidos uniformemente (independientemente de que sean rojos o azules) por el plano al igual que en el dataset original

Submit

You have used 0 of 2 attempts

Pregunta 12

0.0/1.0 point (graded)

Cuantos parámetros aprendió el modelo de Naive Bayes? (sin considerar las priors como parámetros aprendidos)

Cuantos parámetros aprendió el modelo de QDA? (sin considerar las priors como parámetros aprendidos y sumar todos los valores fuera de la diagonal de la matriz de covarianza a pesar de que la matriz sea simétrica))

Submit

You have used 0 of 5 attempts