

Benemérita Universidad Autónoma de Puebla
Facultad de Ciencias de la Computación

Máquinas de aprendizaje

Reporte: Revisión de Naïve Bayes



BUAP

Docente: Abraham Sánchez López

Alumno

Taisen Romero Bañuelos

Matrícula

202055209

Naïve Bayes

Bayes no fue un nombre desconocido para mí cuando leí sus métodos en minería, en este curso y tampoco cuando cursé probabilidad y estadística. Durante la pandemia por Covid-19 descubrí un video en el que se hablaba de las engañosas formas en que uno puede calcular la probabilidad de que suceda un evento (justo como también se relata en el PDF). En su momento no me enteré de muchas de las cosas que se mencionaron en el video, sin embargo, con la lectura de los métodos Bayesianos puedo comparar los ejemplos con los vistos en el video y entender mejor de qué trata el asunto. En el video recuerdo que hablaba de las probabilidades de padecer cáncer, pero para no pensar en fatalidades repasemos lo más importante de los ejemplos de correos no deseados de una dudosa procedencia que seguramente afirma vivir a 3 km de mi casa.

Desde una perspectiva empírica, Naive Bayes no es muy diferente de como nosotros hacemos predicciones sobre eventos futuros, pues, se basa en los resultados de unos datos almacenados (históricos) para determinar que un evento ocurra, con el detalle de que considera a todas las variables disponibles como igual de importantes, cosa que nosotros tal vez no haríamos. Este enfoque, pese a ser razonablemente cuestionado, da buenos resultados (y más si se manejan relativamente pocos datos), y como dice el lema del ingeniero “si funciona no le muevas” pareciera que hay un poco de relajación en cuanto a descubrir a qué se debe este fenómeno (aunque seguramente los científicos más acérrimos se lo toman en serio).

Un talón de Aquiles (cuya solución es bastante sencilla) es que el modelo podría quedar invalidado al tener escenarios nunca vistos (o sea, que numéricamente son un producto por cero). Bueno, la solución es hacer un supuesto matemático para que agregando un valor cualquiera a cada recuento de frecuencias los escenarios ya no sean una multiplicación por cero, y ahora, preferiblemente, se vuelve en un neutro multiplicativo.

Otra ventaja no mencionada de este método Bayesiano es que tiene dificultades cuando hay muchas características numéricas (ya que es un método para clasificación, o sea, datos cualitativos). Su solución es discretizar los datos para poder dividirlos en categorías para reducir la información.

Como dato curioso, este es el video que vi sobre el teorema de Bayes:

<https://www.youtube.com/watch?v=R13BD8qKeTg>