

Aplicación del Clasificador Bayesiano Ingenuo Usando EMV

Febrero del 2025

Estudiante: Javier Horacio Pérez Ricárdez

Asignatura: Aprendizaje Máquina Aplicado

Profesor: Dr. Omar Velázquez López

Actividad: Aplicación del Clasificador Bayesiano Ingenuo

1 Introducción

El *Clasificador Bayesiano Ingenuo* (Naive Bayes, NB) es un modelo de clasificación basado en el teorema de Bayes, que asume que las características son independientes entre sí, dado que se conoce la clase. En este caso, el modelo predice si un paciente es *Sano* o *Enfermo* en base a tres características: *Temperatura*, *Presión Arterial* y *Comorbilidad*.

El código implementa un Naive Bayes utilizando el *Estimador de Máxima Verosimilitud (EMV)*, que es utilizado para estimar los parámetros de las distribuciones de las características. El modelo supone que las características de cada clase siguen una distribución normal (Gaussiana), y los parámetros de dicha distribución (media y varianza) se estiman utilizando el EMV.

2 Funcionamiento del Modelo

El modelo de Naive Bayes con EMV calcula la probabilidad de que un nuevo paciente pertenezca a una clase dada en base a las características proporcionadas. Se asume que las características siguen una distribución normal, y se estiman la media y la varianza de cada clase utilizando los datos de entrenamiento.

2.1 Estimación de Parámetros (EMV)

El *Estimador de Máxima Verosimilitud (EMV)* se usa para estimar la media y la varianza de las características para cada clase. Estas estimaciones se realizan a partir de los datos de entrenamiento.

$$\hat{\mu}_c = \frac{1}{N_c} \sum_{i=1}^{N_c} x_i \quad \text{y} \quad \hat{\sigma}_c^2 = \frac{1}{N_c} \sum_{i=1}^{N_c} (x_i - \hat{\mu}_c)^2$$

donde:

- $\hat{\mu}_c$ es la media de las características para la clase c ,
- $\hat{\sigma}_c^2$ es la varianza de las características para la clase c ,
- N_c es el número de muestras de la clase c ,
- x_i son los valores de las características para la clase c .

2.2 Cálculo de Probabilidades

Una vez que se estiman los parámetros de la distribución, el modelo calcula la probabilidad de que una nueva observación pertenezca a una clase específica utilizando la función de densidad de probabilidad de la distribución normal:

$$P(x|c) = \frac{1}{\sqrt{2\pi\hat{\sigma}_c^2}} \exp\left(-\frac{(x - \hat{\mu}_c)^2}{2\hat{\sigma}_c^2}\right)$$

El modelo luego combina esta probabilidad con las probabilidades a priori de las clases para calcular la probabilidad posterior de cada clase. Finalmente, el modelo selecciona la clase con la mayor probabilidad posterior.

3 Implementación en Código

El código implementa el modelo de Naive Bayes con EMV de la siguiente manera:

```
# Ajuste del modelo (fit)
for c in self.classes:
    X_c = X[y == c]
    self.feature_stats[c] = {
        'mean': X_c.mean(axis=0), # Media de la característica para la clase
        'var': X_c.var(axis=0),    # Varianza de la característica para la clase
    }
```

Aquí, el modelo estima la media y la varianza de las características para cada clase, utilizando el EMV.

3.1 Cálculo de la Probabilidad de Gaussianas

El modelo utiliza la fórmula de la distribución normal para calcular la probabilidad de que una nueva muestra pertenezca a una clase:

```
# Probabilidad de Gaussianas
def _gaussian_prob(self, x, mean, var):
    coeff = 1.0 / np.sqrt(2.0 * np.pi * var)
    exponent = np.exp(-((x - mean) ** 2) / (2 * var))
    return coeff * exponent
```

Esta función calcula la probabilidad de que una característica x siga una distribución normal con media μ y varianza σ^2 .

4 Predicción

Una vez entrenado el modelo, se utiliza para predecir si un paciente es *Sano* o *Enfermo* basándose en sus características:

```
# Entrada del usuario
temperatura = st.number_input("Temperatura (°C)", min_value=34.0, max_value=40.0, value=36.5)
presion = st.number_input("Presión Arterial (mmHg)", min_value=100, max_value=180, value=120)
comorbilidad = st.selectbox("Comorbilidad", label_encoder_comorbilidad.classes_)

# Predicción
nuevo_paciente = np.array([[temperatura, presion, comorbilidad_encoded]])
probas = nb_emv.predict_proba(nuevo_paciente)[0]
prediccion = nb_emv.predict(nuevo_paciente)[0]
clase_predicha = label_encoder_clasificacion.inverse_transform([prediccion])[0]
```

El modelo devuelve la clase predicha (Sano o Enfermo) y las probabilidades asociadas a cada clase.

Predicción con Clasificador Bayesiano Ingenuo (Naive Bayes, NB): Modelo Original y con Ruido en la Temperatura

Febrero del 2025

Estudiante: Javier Horacio Pérez Ricárdez

Asignatura: Aprendizaje Máquina Aplicado

Profesor: Dr. Omar Velázquez López

Actividad: Aplicación del Clasificador Bayesiano Ingenuo

1 Clasificador Bayesiano Ingenuo (Naive Bayes, NB)

En este trabajo se utiliza el Clasificador Bayesiano Ingenuo (Naive Bayes, NB) para la predicción de la condición de salud de un paciente. Este modelo probabilístico se basa en el teorema de Bayes, y asume que las características (en este caso, la temperatura, presión arterial y comorbilidad) son independientes entre sí, lo cual simplifica los cálculos y permite realizar predicciones eficientes.

2 Modelo Original

En este modelo, los datos de entrada son correctos y no contienen información irrelevante ni ruido. La tabla a continuación muestra los datos de entrada utilizados para el modelo con el Clasificador Bayesiano Ingenuo (Naive Bayes):

Temperatura (°C)	Presión arterial	Comorbilidad
36.5	120	Arritmia

Table 1: Datos de entrada del modelo original (sin ruido) utilizando Naive Bayes.

Resultados de la predicción con Naive Bayes:

- **Clase Predicha Original:** Sano (S)
- **Probabilidad de Enfermo (E):** 0.0002
- **Probabilidad de Sano (S):** 0.9998

Temperatura (°C)	Presión arterial	Comorbilidad
40.2	120	Arritmia

Table 2: Datos de entrada con temperatura irrelevante (40.2°C) utilizando Naive Bayes.

3 Modelo Predicción con Ruido o Información Irrelevante

En este modelo, la temperatura se establece en un valor irrelevante (40.2°C), lo que introduce ruido en los datos. La tabla siguiente muestra los datos de entrada con la temperatura irrelevante:

Resultados de la predicción con ruido:

- **Temperatura Introducida:** 40.2°C
- **Comorbilidad Introducida:** Arritmia
- **Clase Predicha con Ruido o Información Irrelevante:** Sano (S)
- **Probabilidad de Enfermo (E):** 0.0000
- **Probabilidad de Sano (S):** 1.0000

4 Análisis de los Resultados

La introducción de la temperatura irrelevante (40.2°C) en el modelo no afecta significativamente la predicción. A pesar de que se introdujo un valor de temperatura que no es relevante para la predicción de la condición de salud, el modelo sigue clasificando a la persona como **Sano (S)** con una probabilidad de 1.0000. Esto indica que el Clasificador Bayesiano Ingenuo (Naive Bayes) no está sensible al valor de la temperatura en este caso particular.

Sin embargo, en situaciones más complejas o con más características irrelevantes, el modelo podría verse afectado. En esos casos, se recomendaría aplicar técnicas de preprocesamiento para filtrar o ajustar las características que no aportan información relevante, o utilizar métodos de selección de características para mejorar la precisión del modelo.

Predicción con Clasificador Bayesiano Ingenuo (Naive Bayes, NB): Modelo Original y con Ruido en la Presión arterial

Febrero del 2025

Estudiante: Javier Horacio Pérez Ricárdez

Asignatura: Aprendizaje Máquina Aplicado

Profesor: Dr. Omar Velázquez López

Actividad: Aplicación del Clasificador Bayesiano Ingenuo

1 Introducción

En este documento se explora el efecto que tiene la introducción de **ruido** o **información irrelevante** en un modelo de **Clasificador Bayesiano Ingenuo (Naive Bayes, NB)**. El análisis se centra en el caso en que una de las características (en este caso, la *presión arterial*) se ve alterada por valores que no corresponden a la realidad o están fuera de la distribución esperada.

2 Efecto del Ruido en la Predicción

Cuando se introduce un **valor de presión arterial muy alto** (240 mmHg), el modelo de Naive Bayes, que asume distribuciones estadísticas de las características, interpreta que este valor es un *outlier* (un valor extremo), lo que impacta directamente en las predicciones. Aquí está el efecto del ruido:

- **Desviación en las probabilidades:** El modelo puede asignar **probabilidades muy bajas** a la clase "Enfermo" (E) cuando la presión arterial está muy lejos de los valores que ha visto durante el entrenamiento. Esto puede ocurrir porque el modelo, al no haber visto presiones tan altas, no sabe cómo interpretar este valor y asume que debe ser más probable que la clase "Sano" (S) se relacione con un valor tan extremo.
- **Impacto en la predicción final:** El modelo predice una clase de manera más sesgada hacia "Sano" debido al **ruido** en la presión arterial. Esto indica que, cuando un atributo clave como la presión arterial se ve alterado por *información irrelevante o incorrecta*, puede llevar al modelo a hacer una predicción **incorrecta**, como en el caso que mencionas con la clase predicha "Sano" cuando podría haber sido "Enfermo".

3 ¿Cómo afecta la presión arterial cuando tiene ruido?

La **presión arterial** es una de las características clave en la predicción del modelo. Si esta característica tiene un valor de *ruido o es irrelevante* (como una presión arterial demasiado alta que no es representativa del conjunto de datos), el modelo puede **no detectar correctamente** patrones relevantes en el conjunto de datos.

En este caso:

- Si el modelo ha aprendido que presiones arteriales **normales** (de 120 a 160 mmHg) están asociadas con ciertas probabilidades de enfermedades (por ejemplo, mayor probabilidad de enfermedad si la presión es más alta), entonces un valor de **240 mmHg** puede ser interpretado de manera errónea.
- Este **ruido** hace que el modelo no se enfoque en lo que realmente es importante (por ejemplo, la temperatura y la comorbilidad), desviando la predicción hacia la clase "Sano", como muestra el resultado donde la probabilidad de "Sano" es de **1.0000**.

4 Conclusión

La **presión arterial** con **ruido o información irrelevante afecta negativamente** la predicción porque puede alterar las probabilidades asignadas a cada clase. Si el modelo se entrena con datos de buena calidad y la presión arterial es una característica clave, valores de ruido pueden hacer que el modelo realice predicciones erróneas. Esto resalta la importancia de **limpiar** y **preprocesar los datos** correctamente antes de entrenar el modelo, especialmente cuando se tiene información que podría no ser representativa o que podría alterar las predicciones de manera no deseada.

Predicción con Clasificador Bayesiano Ingenuo (Naive Bayes, NB): Modelo Original y con Ruido en la Comorbilidad desconocida (Cáncer)

Febrero del 2025

Estudiante: Javier Horacio Pérez Ricárdez

Asignatura: Aprendizaje Máquina Aplicado

Profesor: Dr. Omar Velázquez López

Actividad: Aplicación del Clasificador Bayesiano Ingenuo

Justificación

Cuando el modelo se enfrenta a una **comorbilidad desconocida** (como en este caso '*Cáncer*'), se ve influenciado por la **información irrelevante** que no estaba presente en los datos de entrenamiento. Esto afecta las probabilidades de predicción, lo que se puede observar en las diferencias de los resultados.

El **Clasificador Bayesiano Ingenuo (Naive Bayes, NB)** es un modelo de clasificación probabilística que se basa en el Teorema de Bayes con la suposición de independencia entre las características. Este modelo es muy eficiente y se usa frecuentemente para tareas de clasificación, pero puede verse afectado cuando se le proporcionan características irrelevantes o ruido.

Ejemplo

1. Modelo Original (Datos Correctos):

Temperatura	Presión arterial	Comorbilidad
36.5	120	Ansiedad

Predicción:

- **Clase Predicha:** Sano (S)
- **Probabilidad de Enfermo (E):** 0.0000
- **Probabilidad de Sano (S):** 1.0000

En este caso, el modelo predice correctamente que el paciente está **Sano** con una probabilidad de 100%. El modelo ha aprendido a partir de los datos de entrenamiento y las relaciones entre las características.

2. Modelo con Ruido o Información Irrelevante (Comorbilidad: 'Cáncer'):

Temperatura	Presión arterial	Comorbilidad
38.1	158	Cáncer

Predicción:

- **Clase Predicha:** Sano (S)
- **Probabilidad de Enfermo (E):** 0.0440
- **Probabilidad de Sano (S):** 0.9560

Aquí, al introducir 'Cáncer' como una comorbilidad no presente en el entrenamiento, el modelo sigue prediciendo **Sano** pero con una probabilidad menor (95.6%) para "Sano" y 4.4% para "Enfermo". Esto demuestra cómo un valor irrelevante o desconocido afecta la predicción.

Clasificador Bayesiano Ingenuo (Naive Bayes, NB)

El **Clasificador Bayesiano Ingenuo** es un algoritmo de clasificación que calcula las probabilidades de las clases dado un conjunto de características. Para cada clase, calcula la probabilidad de que las características se presenten en ella, asumiendo que todas las características son independientes entre sí. Este modelo es útil en muchas aplicaciones, pero como hemos visto en el ejemplo anterior, puede ser sensible a la inclusión de información irrelevante, como la comorbilidad 'Cáncer'.

Dado que Naive Bayes hace la suposición de independencia entre las características, cualquier información que no sea relevante o no forme parte del conjunto de entrenamiento puede afectar negativamente la probabilidad final de predicción. En el caso de introducir una comorbilidad como 'Cáncer' que no forma parte del entrenamiento, el modelo puede generar una predicción menos confiable.

Conclusión

El **ruido o la información irrelevante** (como en este caso, una comorbilidad no vista) tiene un impacto en la capacidad del modelo para hacer predicciones precisas. En este caso específico, aunque el modelo sigue prediciendo **Sano** en ambas situaciones, las **probabilidades cambian**, lo que indica que el modelo no está tan seguro como antes.

Impacto en la predicción:

- **Incertidumbre aumentada:** El modelo tiene más dificultad para realizar una predicción precisa cuando se le presenta información que no ha aprendido o que es irrelevante.
- **Reducción de la confianza:** Las probabilidades de la predicción para la clase "Enfermo" (E) aumentan ligeramente, lo que refleja una menor confianza en la predicción.

Respuesta

El impacto de introducir características irrelevantes o con ruido, como una comorbilidad desconocida, es **una disminución de la certeza** en las predicciones del modelo. Aunque la predicción final puede ser correcta (en este caso, "Sano"), las **probabilidades se ven afectadas**, lo que refleja la inseguridad del modelo al enfrentar un valor no conocido.

Este tipo de comportamiento ilustra cómo la presencia de datos **irrelevantes o desconocidos** puede **degradar el rendimiento de un modelo de clasificación** al generar predicciones menos confiables, especialmente cuando la nueva entrada contiene información fuera del dominio entrenado.

Bibliografía (APA)

- Velázquez López, O. (2025). *Apuntes Clasificador Bayesiano Ingenuo*. Universidad Panamericana.
- Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. MIT Press.
- Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.
- Zhang, H. (2004). *The optimality of Naive Bayes*. Proceedings of the Seventeenth International Florida Artificial Intelligence Research Society Conference (FLAIRS 2004), 1-5.