Predicción del Valor de Mercado de Jugadores para PANINOS

1. Predicción del valor de mercado

i) Prueba de significancia de regresión, coeficiente de determinación (R^2) y correlación lineal (r)

- Recta de regresión ($y = \beta 0 + \beta 1 \times x$):

  Donde:

  - y es el valor de mercado del jugador.
  - x es la habilidad de overall.
  - β0 es la ordenada de origen.
  - β1 es la pendiente, que representa cuánto cambia el valor de mercado por cada unidad que aumenta el rendimiento general.

  y^= -15.153.844,42 + 260.857 × overall

  Cada punto que aumente overall, se asocia con un aumento de 260.875 EUR en su valor de mercado.

- Prueba de significancia de regresión:

  Para verificar si la relación entre el rendimiento general y el valor de mercado es significativa, realizamos una prueba de hipótesis para el coeficiente de la pendiente (β1):

  - **Hipótesis nula H0**: β1=0 (no hay relación entre el rendimiento general y el valor de mercado).
  - **Hipótesis alternativa H1**: β1≠0 (hay una relación significativa).

  Como el valor p es muy bajo ($p < 0.05$), rechazamos **H0** y concluimos que overall y el valor de mercado están significativamente relacionados.

- Coeficiente de determinación **R^2**

  En la relación entre habilidad general y valor, $R^2 = 0.386$. Esto indica que aproximadamente el 38.6% de la variabilidad del valor de mercado puede explicarse por la habilidad general.

- Correlación lineal (r):

  r = $\sqrt{R^2}$ = $\sqrt{0.386}$ ≈ 0,621
  La correlación lineal entre overall y el valor de mercado es positiva y moderada, indicando una relación significativa.

ii) Inferencias sobre los parámetros de la recta con un 95% de confianza

Los intervalos de confianza del 95% para los coeficientes de la regresión permiten estimar un rango dentro del cual se espera que caiga el verdadero valor de los coeficientes con un 95% de confianza. Para el análisis:

- Pendiente (β1):

  260.857 $\pm$ 1.96 $\times$ 4.500 = [269.677, 252037]

- Ordenada de origen (β0):

  -15.153.844,42 $\pm$ 1.96 $\times$ 297.000 [-14571724,42, -15735964.42]

iii) Proporción de veces que el valor de mercado supera la incertidumbre de predicción

La incertidumbre de predicción se refiere a la capacidad del modelo para predecir valores futuros. Para un jugador con una característica fija, se puede calcular el intervalo de predicción del valor de mercado y compararlo con la media del valor de mercado.

Intervalo de predicción para un jugador con overall específico:
Supongamos que queremos predecir el valor de mercado para un jugador con un overall de 80:

- Predicción:

  y^= -15.153.844,42 + 260.857 $\times$ 80 = 5.714.715 EUR

- Intervalo de predicción (aproximado):

  [1.125.853,16.408.718] EUR

Esto significa que en la mayoría de los casos (aproximadamente el 95%), los valores de mercado predichos por el modelo se encuentran dentro del intervalo de predicción estimado.

2. Ecuación de predicción del valor de mercado a partir de varias características

i) Método de mínimos cuadrados

Análisis de Indicadores usando Mínimos Cuadrados

Ecuación del Modelo: la ecuación de regresión múltiple para predecir el valor de mercado (Y) a partir de las características X1 (habilidad general), X2 (potencial) y X3 (habilidad de movimientos) es:

Y = B0 + B1X1 + B2X2 + B3X3

Coeficiente de Determinación (R^2):

- En la relación entre habilidad general y valor, $R^2 = 0.386$. Esto indica que aproximadamente el 38.6% de la variabilidad del valor de mercado puede explicarse por la habilidad general. Es una relación moderada.
- Para la relación entre potencial y valor, $R^2 = 0.286$, lo que sugiere que el potencial explica alrededor del 28.6% de la variabilidad del valor. Es una relación más débil en comparación con la habilidad general.
- Para movimientos y valor, $R^2 = 0.022$, indicando que solo el 2.2% de la variabilidad del valor se explica por los movimientos, lo que es muy bajo y sugiere que esta variable tiene poca relevancia.
- $R^2 = 0.437$: esto indica que aproximadamente el 43.7% de la variabilidad del valor de mercado (Y) se puede explicar por las características X1, X2, X3. Este valor sugiere que el modelo tiene una capacidad moderada para predecir el valor basado en las características dadas.

Correlación (r):

- La correlación entre habilidad general y valor es r(x1) = 0.621, lo que indica una correlación positiva moderada. A mayor habilidad, mayor valor.
- Para el potencial, r(x2) = 0.534, lo que también muestra una correlación positiva moderada, aunque menor que con la habilidad general.
- La correlación entre movimientos y valor es r(x3) = 0.148, lo que indica una correlación débil.
- r(x1, x2, x3) = 0.661: esto implica una correlación positiva moderada entre las características y el valor de mercado. A medida que las características aumentan, el valor de mercado tiende a aumentar también.

Los coeficientes B1, B2 y B3 indican la contribución de cada variable x1 (habilidad general), x2 (potencial), y x3 (habilidad de movimientos) al valor de mercado:

- B1 ≈ 194.146,68: por cada unidad de aumento en x1, el valor de mercado se incrementa en 194.146,68.
- B2 ≈ 131.309,30: cada unidad adicional de x2 aumenta el valor en 131.309,30.
- B3 ≈ 99912,08: un aumento en x3 eleva el valor en 99.912,08.

ii) Método de Descenso por Gradiente

Comparación entre descenso por gradiente y ecuaciones normales.

- **Descenso por gradiente**:

➔ Parámetros:

θ(dg) = [-16722,35, 22719,08]

➔ Últimos valores del costo:

[3733217856721.50,3733217565503.86,3733217274286.29,3733216983068.78,3733216691851.34]

➔ Tasa de aprendizaje (alpha): 0.00001.

➔ Número de iteraciones: 10000 (se detuvo antes de completarlas por convergencia).

- **Ecuaciones normales:**

➔ Parámetros:

θ(normal) = [−15153844.42,258872.59]

Comparación: Los parámetros obtenidos mediante el descenso por gradiente y las ecuaciones normales son diferentes. Esto puede deberse a que el descenso por gradiente no ha convergido a una solución óptima o que hay un problema en la escala de los datos que afecta la convergencia.

ii) Interpretación del criterio de corte

El criterio de corte en el algoritmo de descenso por gradiente se basa en detener el proceso si la diferencia en el costo entre iteraciones consecutivas es menor que un umbral específico (1e-6).

Posibles fallas:

- **Convergencia prematura**: Si el umbral es demasiado bajo, el algoritmo puede detenerse antes de alcanzar el mínimo global, especialmente en regiones donde la función de costo es casi plana.

- **Sensibilidad al ruido**: Si hay ruido en los datos, el costo podría fluctuar, lo que podría llevar a un resultado engañoso.

Alternativa sugerida: Un enfoque más robusto podría ser implementar un número fijo de iteraciones o considerar la mejora del costo en varias iteraciones antes de detenerse. Por ejemplo, podrías observar el costo en las últimas 10 iteraciones y detenerte solo si no hay mejora significativa en ese rango.

3. Comportamiento del método de descenso por gradiente

El método de descenso por gradiente **no siempre garantiza la convergencia al mínimo global**. La convergencia depende de la naturaleza de la función objetivo:

- **Función no convexa**: Si la función es no convexa, el descenso por gradiente puede converger a un mínimo local en lugar del mínimo global. Esto es especialmente relevante en problemas con múltiples valles y picos.

Contraejemplo. Consideremos la función:

$$f(x) = x^4 - 4x^2 + 4$$

Esta función tiene un mínimo global en $x=0$ y dos mínimos locales en $x = \pm 2$. Si inicias el descenso por gradiente en un punto cercano a $x=2$, es posible que el algoritmo converja a ese mínimo local y no al mínimo global en $x=0$.

Resumen

1. Los resultados del descenso por gradiente y las ecuaciones normales no coinciden, lo que puede indicar problemas de convergencia.
2. El criterio de corte actual puede resultar en una convergencia prematura; se sugiere un enfoque alternativo.
3. El descenso por gradiente no siempre garantiza convergencia al mínimo global, especialmente en funciones no convexas, lo que puede ilustrarse con un contraejemplo.

Cálculos, gráficos y explicaciones:

https://colab.research.google.com/drive/1rS7rTXaI7_zfS425N-H4SIskDzbJl9AN#scrollTo=dNAijB8DqnqR