Matematica IV

Autores: Ortega Mariano, Daguerre Elgarte Ivy
Código Auxiliar: Código Proyecto

Parte 1: Predicción del valor de mercado

a) Recta de regresión para predecir el valor de mercado de un jugador a partir de la característica más relevante (a la que se destinará mayor proporción del presupuesto), respaldada por:

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

ii) Inferencias sobre los parámetros de la recta, estimando las fluctuaciones con una confianza del 95%.

iii) La proporción de veces que el valor de mercado supera la incertidumbre de predicción comparada con la respuesta media del valor de mercado para una característica fija, ambas con la misma confianza y ancho mínimo

i)

Para encontrar la recta que mejor se ajusta a los datos proporcionados permitiendo así predecir el valor de mercado de un jugador, utilizaremos regresión lineal simple. Queremos predecir el valor de mercado basándonos en una única característica clave, aquella a la que destinamos la mayor parte del presupuesto.

Teniendo un punto $(x_i,\ y_i)$ la distancia vertical entre este punto y la recta de regresión que estamos buscando representa el error de predicción para ese jugador específico. Si elevamos al cuadrado cada una de estas desviaciones verticales, así siempre son positivas, obtenemos una función $f(b_0,\ b_1)$ que representa el error total de la recta con respecto a los puntos. El objetivo es obtener los estimadores de b_1 y b_0 que reducen al mínimo la $f(b_0, b_1)$. Entonces la línea de regresión estimada tiene la ecuación $\widehat{y} = \widehat{\beta}_1 x + \widehat{\beta}_0$

El valor de los estimadores se obtiene con la derivada parcial de la función de error con respecto a cada una, hallando las ecuaciones normales y resolviendo el sistema de ecuaciones para β_1, quedando

$$\widehat{\beta}_0 = \bar{y} - \widehat{\beta}_1 * \bar{x}\ ;\ \hat{\beta}_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

Para todas las características dadas los estimadores son:

x_1 = attacking_crossing
x_2 = attacking_finishing
x_3 = attacking_heading_accuracy

x_4 = attacking_short_passing
x_5 = attacking_volleys
x_6 = skill_dribbling
x_7 = skill_curve
x_8 = skill_fk_accuracy
x_9 = skill_long_passing
x_{10} = skill_ball_control
x_{11} = movement_acceleration
x_{12} = movement_sprint_speed
x_{13} = movement_agility
x_{14} = movement_reactions
x_{15} = movement_balance
x_{16} = power_shot_power
x_{17} = power_jumping
x_{18} = power_stamina
x_{19} = power_strength
x_{20} = power_long_shots
x_{21} = mentality_aggression
x_{22} = mentality_interceptions
x_{23} = mentality_positioning
x_{24} = mentality_vision
x_{25} = mentality_penalties
x_{26} = mentality_composure
x_{27} = defending_standing_tackle
x_{28} = defending_sliding_tackle
x_{29} = goalkeeping_diving
x_{30} = goalkeeping_handling
x_{31} = goalkeeping_kicking
x_{32} = goalkeeping_positioning
x_{33} = goalkeeping_reflexes

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}

$\widehat{\beta}_1$	73285.618118	63064.59341593000135617331	53646.95345039999665459618	115433.98028557999350596219	78064.60213759999896865338	73124.57617765999748371542	81358.80629038000188302249	74313.54165785999794024974	104638.44507051999971736223	92367.78040709999913815409
$\widehat{\beta}_0$	-1411095.74312713998369872570	-663298.07477382000070065260	-558099.76346842001657932997	-4551790.19316270016133785248	-1106168.33694540988653898239	-1836993.66021140990778803825	-1614288.20121809002012014389	-922822.59599637996871024370	-3284717.89387637982144951820	-3177112.34044871991500258446

	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}	x_{19}	x_{20}
$\widehat{\beta}_1$	60831.89307312999881105497	63286.49239240000315476209	69847.63034258999687153846	312472.75526110001374036074	46960.58648030000040307641	135579.79982029998791404068	60619.49563882999791530892	72776.19012925999413710088	59151.23845692999748280272	71058.02279086000635288656
$\widehat{\beta}_0$	-1685771.24823908996768295765	-1846184.53947828989475965500	-2198860.13106798008084297180	-17026454.35890698060393333435	-776839.27539876999799162149	-5605299.46207770984619855881	-1690632.05881301988847553730	-2330686.54998341016471385956	-1604810.99319284991361200809	-1097771.93159305001609027386

	x_{21}	x_{22}	x_{23}	x_{24}	x_{25}	x_{26}	x_{27}	x_{28}	x_{29}	x_{30}
$\widehat{\beta}_1$	60771.26598984000156633556	38846.77615058000083081424	66611.01039618000504560769	136227.95652646999224089086	72567.79070379999757278711	188242.58163093999610282481	31401.75156490000154008158	27193.18163718000141670927	-3500.46274356999992960482	-3276.07038813000008303788
$\widehat{\beta}_0$	-1147315.09472492011263966560	428149.89484313997672870755	-1122776.76328359008766710758	-5108096.56276862975209951401	-1262098.93056269991211593151	-8689241.97894736006855964661	730662.45558841002639383078	986258.89508289995137602091	2282382.08243290008977055550	2278005.1643931

	x_{31}	x_{32}	x_{33}
$\widehat{\beta}_1$	-3969.93323905	-2891.77044992	-3000.0607193
$\widehat{\beta}_0$	2288742.54506815	2271735.10661679	2274468.22415621

Teniendo la recta de regresión para cada una ahora hay que elegir cuál es la que mejor se ajusta a los datos

Las desviaciones verticales entre los valores observados y los estimados se llaman residuos y la suma al cuadrado de estos es la desviación no explicada o SCE. Por otro lado se tiene la variabilidad explicada SCR y la suma de ambas es la variación total denotada como $STC(S_{yy})$. Reordenando la igualdad $STC = SCR + SCE$ y dividiendo a ambos miembros por la variabilidad total se obtiene la proporción de variación observada que puede ser explicada por el modelo de regresión lineal simple, llamado **coeficiente de determinación**, denotado por .

$$R^2 = 1 - \frac{SCE}{STC}$$

El coeficiente de determinación siempre es menor o igual a 1 y mayor o igual a 0. Entre más cerca de 1 se encuentre mejor es el modelo al explicar la variación de Y

Por otro lado tenemos el coeficiente de correlación lineal, denotado con r, que es la medida que podemos usar para describir que tan fuerte es la relación lineal entre las variables x e y. Si existe una relación inversa (esto es, si y disminuye al aumentar x), entonces r tomará valores entre 0 y -1, de manera similar, si existe una relación directa, si y aumenta al aumentar x, entonces r obtendrá un valor entre 0 y 1. Se puede calcular con 3 formulas:

$$r = \frac{Cov(x,y)}{\sqrt{S_{xx}S_{yy}}} \quad r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} \quad r = \sqrt{R^2}$$

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
R^2	0.06797879	0.05863771	0.03316713	0.10868381	0.07301351	0.07249084	0.08450379	0.06304123	0.09723168	0.09027988
r	0.26072742	0.24215225	0.18211846	0.32967228	0.27021012	0.26924122	0.29069535	0.25108012	0.31181996	0.30046611

	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}	x_{19}	x_{20}
R^2	0.03166739	0.03319463	0.04008138	0.31144334	0.01678083	0.12546648	0.01994865	0.05129204	0.02106479	0.07227973
r	0.17795333	0.18219393	0.20020335	0.55807109	0.12954084	0.35421248	0.14123971	0.22647745	0.14513713	0.2688489

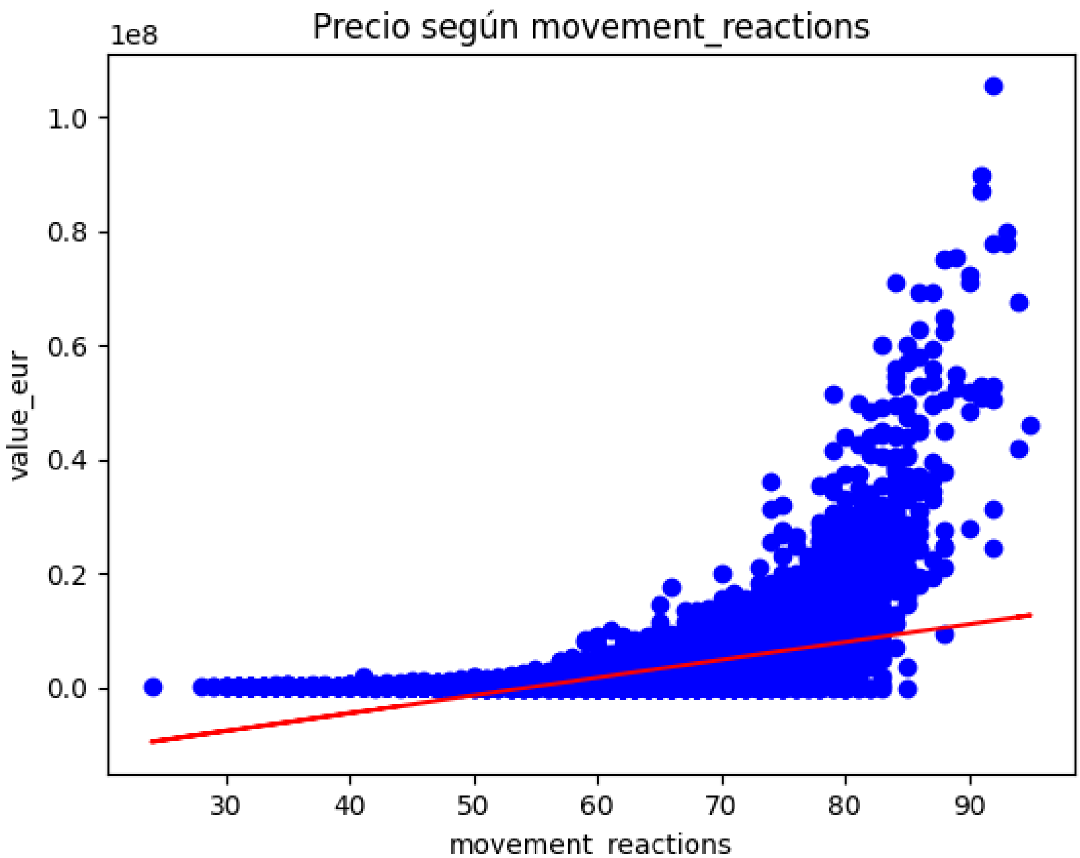
	x_{21}	x_{22}	x_{23}	x_{24}	x_{25}	x_{26}	x_{27}	x_{28}	x_{29}	x_{30}
R^2	0.04198011	0.02492783	0.06442997	0.13425635	0.04967725	0.1998766	0.01734888	0.01247068	0.00014541	0.00011698

r	0.20489049	0.1578855	0.25383059	0.36641008	0.22288394	0.44707561	0.13171516	0.1116722	-0.01205859	-0.0108157

	x_{31}	x_{32}	x_{33}
R^2	0.00016519	9.30138042e-05	0.00011049
r	-0.01285274	-0.00964437	-0.01051163

Con los resultados obtenidos se puede observar que si bien ninguna tiene una correlación lineal fuerte con la propiedad que se intenta predecir, la característica más relevante y que mejor explica el valor de mercado del jugador es “movement_reactions”, ya que tiene tanto el mayor coeficiente de determinación como el mayor coeficiente de correlación lineal

La recta de regresión lineal que podemos dibujar utilizando los estimadores que calculamos anteriormente en base a los valores de “movement_reactions” queda como



Test de significancia

El test de significancia sirve para evaluar si la relación entre la variable predictora y la variable de respuesta es estadísticamente significativa. Se realiza sobre la pendiente β_1, tomando H_0: $\beta_1 = 0$ y H_1: $\beta_1 \neq 0$, y utilizando el estadístico de prueba

$$T = \frac{\widehat{\beta}_1 - \theta}{\sqrt{\frac{\widehat{\sigma}^2}{S_{xx}}}} \sim t_{n-2} \text{ bajo } H_0$$

donde

$$\widehat{\sigma}^2 = \frac{SCE}{n-2} \text{ y } S_{xx} = \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}$$

Aceptar H_0 es equivalente a concluir que no hay ninguna relación lineal entre x e y, por otro lado si el resultado es el contrario, implica que x tiene importancia al explicar la variabilidad de y. Además, como no se tiene un nivel de significancia α, se debe usar el P-valor, que en este caso es igual a $P(|T| > |t_0|)$.

Reemplazando los valores nos queda

$$t_0 = \frac{312472.7552611 - 0}{\sqrt{\frac{1.79277691e+13}{1573139.15619724}}} = 2.6131$$

Dado este valor si utilizamos la aplicación “Probability Distributions” con la distribución de Student n-2 grados de libertad (n siendo 18944) el P-valor da un resultado de 0

Como $0 < 0.05$ hay evidencia suficiente para rechazar H_0 y concluir que la característica “movement_reactions” tiene relevancia en el valor de mercado del jugador

ii)

Para este inciso hay que calcular los intervalos de confianza del 95% para los valores de β_1 y β_0, es decir utilizaremos un $\alpha = 0.05$

Para β_1 tenemos el mismo estadístico usado para el test de significancia

$$P(-t_{\frac{\alpha}{2},n-2} < \frac{\hat{\beta}_1 - \theta}{\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}} < t_{\frac{\alpha}{2},n-2})$$

y el intervalo es

$$\left[\hat{\beta}_1 - t_{\frac{\alpha}{2},n-2}\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} \;;\; \hat{\beta}_1 + t_{\frac{\alpha}{2},n-2}\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}\right]$$

$$\left[312472.7552 - 1.9604\sqrt{\frac{1.79277691e+13}{1573139.15619724}} \;;\; 312472.7552 + 1.9604\sqrt{\frac{1.79277691e+13}{1573139.15619724}}\right]$$

$$[305{,}851.32 \;;\; 319{,}094.20]$$

Para β_0 tenemos el estadístico

$$\left[\hat{\beta}_0 - t_{\frac{\alpha}{2},n-2}\sqrt{\hat{\sigma}^2\left(\frac{1}{n}+\frac{\bar{x}^2}{S_{xx}}\right)}\ ;\hat{\beta}_0 + t_{\frac{\alpha}{2},n-2}\sqrt{\hat{\sigma}^2\left(\frac{1}{n}+\frac{\bar{x}^2}{S_{xx}}\right)}\right]$$

$$\left[-17026454.35890698 - 1.9604\sqrt{1.79277691e+13\left(\frac{1}{18944}+\frac{61.60942778716216^2}{1573139.15619724}\right)}\ ;\ -17026454.35890698 + 1.9604\sqrt{1.79277691e+13\left(\frac{1}{18944}+\frac{61.60942778716216^2}{1573139.15619724}\right)}\right]$$

$$[-17438535.04157005;\ -16614373.6762439]$$

iii)

Para hallar la proporción de veces que el valor de mercado supera la incertidumbre de predicción comparada con la respuesta media del valor de mercado para una característica fija, hay que calcular haciendo el error del intervalo para la media dividido el error del intervalo de predicción

Los intervalos de confianza tienen forma $[\widehat{\Theta} - \varepsilon; \widehat{\Theta} + \varepsilon]$, donde $\widehat{\Theta}$ es el estimador y ε es el error. El cociente de ambos errores nos queda como

$$\frac{t_{\frac{\alpha}{2},n-2}\sqrt{\hat{\sigma}^2\left(1+\frac{1}{n}+\frac{(x^*-\bar{x})^2}{S_{xx}}\right)}}{t_{\frac{\alpha}{2},n-2}\sqrt{\hat{\sigma}^2\left(\frac{1}{n}+\frac{(x^*-\bar{x})^2}{S_{xx}}\right)}}$$

El ancho del intervalo de confianza depende de $(x^* - \bar{x})$, por lo tanto este será mínimo cuando el valor de x* se acerque al valor de $\bar{x}$ y crece a medida que x* se aleje del promedio. Como lo que queremos es un ancho mínimo, igualamos x* a $\bar{x}$ la resta nos queda igual a cero, anulando todo su término. Haciendo eso y cancelando de ambas partes ya que se usa el mismo nivel de confianza, queda

$$\frac{\sqrt{\hat{\sigma}^2\left(1+\frac{1}{n}\right)}}{\sqrt{\hat{\sigma}^2\frac{1}{n}}} = \frac{\sqrt{\hat{\sigma}^2}\sqrt{\left(1+\frac{1}{n}\right)}}{\sqrt{\hat{\sigma}^2}\sqrt{\frac{1}{n}}} = \frac{\sqrt{\left(1+\frac{1}{n}\right)}}{\sqrt{\frac{1}{n}}} = \frac{\sqrt{\frac{1}{n}}\sqrt{n+1}}{\sqrt{\frac{1}{n}}} = \sqrt{n+1}$$

reemplazando n en la fórmula

$$\sqrt{18944 + 1} = \sqrt{18945} = 137.64083$$

Con lo que concluimos que el valor de mercado supera la incertidumbre de predicción algo más de 137 veces más frecuentemente de lo que supera a la respuesta media, para la característica “movement reactions”.

Parte 2: Ecuación de predicción del valor de mercado

Para esta sección definiremos las siguientes variables relevantes, en base a las cuales realizaremos la regresión lineal múltiple $y = b0 + b1x_1 + b2x_2 + b3x_3$.

y = value_eur
x_1 = movement_reactions
x_2 = mentality_composure
x_3 = mentality_vision

Con el fin de calcular los coeficientes $b0$, $b1$, $b2$, y $b3$ mediante el método de mínimos cuadrados, buscaremos minimizar la función:

$$f(b0,\ b1,\ b2,\ b3) = \sum_{i=1}^{n} (y_i - \widehat{y}_i)^2$$

Para ello desarrollamos $\widehat{y}$, calculamos sus 4 derivadas parciales, las igualamos a 0, y pasando algunos términos obtenemos el siguiente sistema de ecuaciones:

$$\sum y_i = nb_0 + b_1 \sum x_{1i} + b_2 \sum x_{2i} + b_3 \sum x_{3i}$$

$$\sum y_i x_{1i} = b_0 \sum x_{1i} + b_1 \sum x_{1i} x_{1i} + b_2 \sum x_{1i} x_{2i} + b_3 \sum x_{1i} x_{3i}$$

$$\sum y_i x_{2i} = b_0 \sum x_{2i} + b_1 \sum x_{2i} x_{1i} + b_2 \sum x_{2i} x_{2i} + b_3 \sum x_{2i} x_{3i}$$

$$\sum y_i x_{3i} = b_0 \sum x_{3i} + b_1 \sum x_{3i} x_{1i} + b_2 \sum x_{3i} x_{2i} + b_3 \sum x_{3i} x_{3i}$$

Reemplazando:

$$42146863000 = 18944b_0 + b_1 1167129 + b_2 1098348 + b_3 1019722$$

$$3088207239000 = b_0 1167129 + b_1 73479289 + b_2 69109904 + b_3 64055331$$

$$2967287763000 = b_0 1098348 + b_1 69109904 + b_2 66462642 + b_3 61223055$$

$$2754741110000 = b_0 1019722 + b_1 64055331 + b_2 61223055 + b_3 58457748$$

Luego sabemos que podemos expresar esta igualdad como una multiplicación de matrices:

$$\underbrace{\begin{bmatrix} 18944 & 1167129 & 1098348 & 1019722 \\ 1167129 & 73479289 & 69109904 & 64055331 \\ 1098348 & 69109904 & 66462642 & 61223055 \\ 1019722 & 64055331 & 61223055 & 58457748 \end{bmatrix}}_{X} \bullet \underbrace{\begin{bmatrix} b0 \\ b1 \\ b2 \\ b3 \end{bmatrix}}_{b} = \underbrace{\begin{bmatrix} 42146863000 \\ 3088207239000 \\ 2967287763000 \\ 2754741110000 \end{bmatrix}}_{Y}$$

Siendo que la matriz X es siempre cuadrada por definición, con $\mathrm{k} + 1$ filas y $\mathrm{k} + 1$ columnas, siendo k el número de variables que influencian a Y, este sistema es fácilmente resoluble calculando la inversa de la matriz X, X^{-1}, y luego multiplicando de ambos lados por esta inversa. El nuevo sistema de matrices dicta que $X^{-1}Xb = X^{-1}Y$. Sabemos también que XX^{-1} es igual a la matriz identidad de tamaño $\mathrm{k} + 1$, y por definición de la identidad hace de neutro en la multiplicación con b.

Tras simplificar, la última iteración de nuestro sistema de matrices declara la igualdad entre b y $X^{-1}Y$, que son dos matrices de valores conocidos:

$$\underbrace{\begin{bmatrix} 0.002478 & -0.00003779 & 0.0000005186 & -0.000002353 \\ -0.00003779 & 0.000001225 & -0.0000005683 & -0.00000008802 \\ 0.0000005186 & -0.0000005683 & 0.000000911 & -0.0000003403 \\ -0.000002353 & -0.00000008802 & -0.0000003403 & 0.0000005111 \end{bmatrix}}_{X} \bullet \underbrace{\begin{bmatrix} 42146863000 \\ 3088207239000 \\ 2967287763000 \\ 2754741110000 \end{bmatrix}}_{Y} = \underbrace{\begin{bmatrix} b0 \\ b1 \\ b2 \\ b3 \end{bmatrix}}_{b}$$

Resolviendo obtenemos los coeficientes:

$$
\underset{XY}{\begin{bmatrix} -\ 17227131.92 \\ 261885.52 \\ 32244.70 \\ 26897.10 \end{bmatrix}} = \underset{b}{\begin{bmatrix} b0 \\ b1 \\ b2 \\ b3 \end{bmatrix}}
$$

Con el sistema de matrices resuelto, podemos trazar la recta de regresión lineal de value_eur en base a nuestras tres estadísticas de la siguiente forma.

$$\hat{y} = -\ 17227131.92 + 261885.52x_1 + 32244.70x_2 + 26897.10x_3$$

i) Teniendo calculados los coeficientes $b0$ a $b3$, buscaremos obtener los coeficientes de determinación y la correlación del modelo.

$$SCE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

Calculamos el SCE, o suma de cuadrados de error, utilizando software para procesar los datos del archivo players.csv, obteniendo el resultado $SCE = 3.3449575127625146 \times 10^{17}$.

Acto seguido calculamos el STC, definido como $\sum_{i=1}^{n} (y_i - \bar{y})^2$, que representa la desviación total de la media, y cuyo resultado da $4.931878872786333 \times 10^{17}$.

Finalmente, calculamos R^2 calculando la diferencia entre 1 y el cociente de SCE sobre STC.

$$R^2 = 1 - \frac{SCE}{STC} = 1 - \frac{3.3449575127625146 \times 10^{17}}{4.931878872786333 \times 10^{17}} = 0.321768113320931$$

$$r = \sqrt{R^2} = \sqrt{0.32176811332093} = 0.5672460782772597$$

Para la raíz de R^2 tomamos solo la raíz positiva, ya que sabiendo que los coeficientes de todas las variables involucradas en la regresión lineal múltiple son positivos, podemos afirmar que la correlación que haya será positiva también, puesto que en este esquema cada punto adicional puesto en cualquier x_k agrega una fracción positiva de su valor a y.

Interpretando los coeficientes podemos afirmar que si bien las predicciones no son despreciables, ya que se pudo predecir aproximadamente el 32% de la variación del precio del jugador con tan solo 3 de las múltiples variables involucradas, este sigue sin ser un buen modelo de predicción. El coeficiente de determinación se halla demasiado lejano a la unidad como para que podamos acercarnos a realmente predecir el precio de jugadores contando con tan poca información.

Del lado de r podemos afirmar que la correlación entre nuestras variables y el valor del jugador es positiva, es decir que esperamos que estas cantidades crezcan juntas, pero al ser un valor tan lejano a la unidad, no podemos esperar demasiada consistencia entre sus aumentos, y podría crecer significativamente sin que esto implique que los valores de nuestras variables estén aumentando, y lo reverso también es válido.

Finalmente, y si bien no es pedido por el inciso, consideramos importante recalcar que este modelo busca descubrir correlación, y no causalidad, y que supondría un salto lógico suponer que sí se halló una correlación fuerte entre una o varias características y el precio de un jugador, aumentar dichas características aumentaría también el precio. Podrían haber muchas explicaciones de porque estas variables crecen, o decrecen, juntas, podría ser que el precio determine las estadísticas, y no las estadísticas al precio, o que tanto las estadísticas como el precio sean causadas por algún factor externo que no está siendo tenido en cuenta.

ii) Si bien el método es distinto, y los coeficientes no son idénticos a los obtenidos mediante la técnica de mínimos cuadrados, los resultados por descenso de gradiente son ampliamente similares:

$$b0 = -17227131.44735364$$

$$b1 = 261885.51154327$$

$$b2 = 32244.70231416$$

$$b3 = 26897.09727634$$

Los parámetros utilizados para esta instancia de descenso de gradiente fueron $\alpha = 0.00001$, tolerancia = 0.000001, punto de inicio = [0,0,0,0], e iteraciones = 44558731. Lamentablemente, por la magnitud de datos a procesar, no fue posible procesar todos los datos en una única ejecución del algoritmo de descenso de gradiente, para solventar esta complicación, los datos fueron calculados usando no uno sino dos descensos de gradiente, el primero descendió hasta el punto [-17212328.15788638, 261659.65930791, 32247.8544781, 26883.04967699], desde el cual se continúo el descenso con los mismos parámetros antes mencionados, a excepción de tener ahora $\alpha = 0.0001$, y contar con el actualizado punto de inicio. El conteo de iteraciones toma en cuenta las iteraciones de ambos descensos, y plantea su suma.

Finalmente, presentamos el listado de las últimas iteraciones del segundo descenso de gradiente, con paso de 0.0001:

[-17227131.44734863, 261885.51154319, 32244.70231416, 26897.09727633]

[-17227131.44734963, 261885.51154321, 32244.70231416, 26897.09727633]

[-17227131.44735064, 261885.51154322, 32244.70231416, 26897.09727633]

[-17227131.44735164, 261885.51154324, 32244.70231416, 26897.09727634]

[-17227131.44735264, 261885.51154325, 32244.70231416, 26897.09727634]

[-17227131.44735364, 261885.51154327, 32244.70231416, 26897.09727634]

iii) Los algoritmos de descenso de gradiente son varios, hay multitud de implementaciones distintas, y no todas comparten condición de corte, sin embargo, la condición de corte más universal de este algoritmo es la de detención por gradiente pequeño, esto es, cuando el algoritmo se detiene porque detecta que la actualización de las variables del gradiente están cambiando en magnitudes muy pequeñas. Al detectar el algoritmo que las variables están cambiando muy poco asume que se hallan cerca de su valor mínimo, y detiene la búsqueda, mas esta es una falla del algoritmo, pues podría ser el caso que la velocidad de actualización se reduzca por haber llegado a un mínimo local no global, o a un punto silla. Cuando el algoritmo se detiene por alguno de esos motivos, el algoritmo falla en su misión de obtener los valores que minimizan la función provista.

Lamentablemente no tenemos la formación para ofrecer una reforma de valor al algoritmo, pero creemos que la implementación de Descenso de Gradiente con Momento podría conformar una buena modificación al original para arreglar la problemática mencionada. Este tipo de Descenso de Gradiente registra una dirección promedio en base a iteraciones anteriores del algoritmo, y así evita depender exclusivamente del gradiente correspondiente al instante actual. El momento acumulado le facilita a esta versión del Descenso de Gradiente atravesar algunos puntos silla, mínimos locales, y "valles" largos con valores bajos.

Si bien esta solución no es perfecta, desconocemos una que lo sea, y al cambiar la forma de cálculo del gradiente, que es parte de cómo el algoritmo juzga su condición su corte, por proxy modifica su condición de corte, haciéndola menos propensa a fallar.

Parte 3: Comportamiento del método de descenso por gradiente

Como fue destacado en el inciso iii) de la Parte 2, no, lamentablemente no siempre converge al mínimo de la función, y esto se debe a que la condición de corte depende de una ralentización de la velocidad descenso, y esta puede ser causada por puntos silla y mínimos locales, que pueden llegar a engañar a la función, la cual se detendrá como si hubiera llegado al mínimo global, aunque en realidad sean otras las causas de la ralentización de la actualización del gradiente.

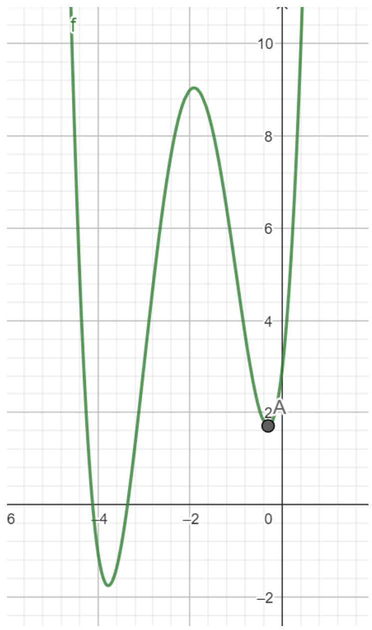
<u>Contraejemplo</u>

Consideremos la función $f(x) = ((x + 2)^2 - 3)^2 + x + 2$. Esta función tiene múltiples mínimos locales y un mínimo global. Si inicializamos el descenso por gradiente en un punto cercano a un mínimo local pero lejos del mínimo global, el algoritmo puede quedarse atrapado en ese mínimo local. Esto se debe a que, conforme se acerca a un mínimo local, el gradiente tiende a reducirse y las actualizaciones se hacen más pequeñas, llevando al algoritmo a detenerse prematuramente.

Además, con una tasa de aprendizaje inadecuada (demasiado baja o demasiado alta), el algoritmo puede reducir su velocidad de descenso y confundir un mínimo local con el mínimo global. Esto ilustra uno de los principales problemas del método de descenso por gradiente.

Decidimos utilizar un contraejemplo con una función sencilla del formato y = f(x) por motivos de simplicidad, y para resolverla utilizamos otro programa de descenso de gradiente, distinto al utilizado para los cálculos de la parte dos. La motivación detrás de esto es que el primer código estaba diseñado para regresión lineal múltiple, y siempre admitía al intercepto como valor a calcular. Dado que aquí queremos calcular un solo x, consideramos apropiado el uso de un algoritmo modificado, sin embargo consideramos esencial enfatizar que ambos algoritmos son algoritmos de Descenso de Gradiente, comparten condición de corte, utilizan el concepto de salto, y actualizan sus mínimos estimados en base a la dirección y velocidad de descenso, es decir, funcionan de forma análoga, más allá de que estimen el gradiente de forma distinta, y todo lo demostrado bajo este ejemplo usando este algoritmo aplica al descenso de gradiente en general, ya que respecta a la condición de corte del mismo, que es universal.

Visualización de la función, con su mínimo estimado:



Parámetros y resultados de ejecución: $\alpha = 0.001$, tolerancia = 0.0001, valor de inicio = 0, número de iteraciones = 485, x estimado =-0.31121975811049024.

Últimas ejecuciones del algoritmo:

-0.31121900491507076
-0.3112191199030829
-0.31121923233562754
-0.31121934226949743
-0.311219449760223
-0.3112195548621005
-0.311219657628219
-0.31121975811049024