

Deporte

¿Se puede predecir quién ganará con base a patrones estadísticos?

Una de las grandes ventajas evolutivas del ser humano y que nos ha permitido sobrevivir hasta el día de hoy es la capacidad de predicción. Existen contextos donde la predicción puede ser usada hasta para conservar la vida, por ejemplo, si se escucha el ruido de un automóvil cuando se está cruzando la calle, se aprieta el paso para evitar un accidente. Si se percibe un objeto que va directo a la cabeza, posiblemente el arco reflejo haga mover la cabeza en fracciones de milisegundos. Estas predicciones son a corto plazo y están integradas como heurísticas en nuestro cerebro; es decir, se actúan sin alcanzar a analizarlas. Existen otro tipo de predicciones que también son de interés, más aún cuando existen acciones que podemos tomar para obtener una ganancia del evento que sucederá. El caso clásico es sin lugar a dudas las apuestas deportivas y qué mejor evento que uno que contemple la mayor cantidad de deportes, los juegos olímpicos. Pero, ¿se puede predecir quién ganará?, ¿qué se debe tomar en cuenta?, ¿qué tan seguros podemos estar de esa predicción?

Para comenzar con este breve análisis, presentemos algunos datos. El próximo 5 de agosto darán inicio los XXXI Juegos Olímpicos en Río de Janeiro, Brasil. En estos juegos participarán 10 500 atletas de 206 países en 42 deportes, los cuales se organizarán en 306 eventos deportivos. Y así podemos ir incluyendo más cifras de una organización monumental en cuanto al tamaño de las variables que intervienen. Pero considerando todo este contexto, ¿qué otro dato se puede usar para predecir?... ¡Acertaste! ¡Los comportamientos históricos de los contendientes!

¿Cómo podemos tomar en cuenta estos comportamientos históricos? Pues bien, podemos codificarlos en variables numéricas. Tomemos el fútbol como caso de prueba. Para predecir cuál de 2 equipos que se enfrentan ganará se deben analizar variables que nos digan algo de su fortaleza, por ejemplo: en cuál zona se clasificó al evento, cuántos goles mete en promedio, cuántos goles le meten en promedio, qué tan común es que le expulsen jugadores, cómo juega de local, cómo juega de visitante, qué experiencia profesional en años tienen sus jugadores, en qué ligas importantes militan sus jugadores y así podemos ir re-



colectando información, la cual nos dará una fotografía completa del equipo. Pero falta algo más al modelo, ya que se enfrentan 2 equipos, también se pueden analizar el comportamiento histórico entre ambos equipos, o contra equipos de la misma zona.

Son muchas las variables, ¿cuáles son las más importantes?, ¿cómo le voy a hacer para decidir cuando tengo 80 variables por cada equipo? ¡Excelente pregunta! Estas variables, como mencionamos al inicio las codificaremos. Por ejemplo, para representar la potencia goleadora del equipo lo que haremos es calcular la razón entre el número de goles que anota y le anotan, es decir, supongamos que el equipo cuando ha participado en juegos olímpicos ha anotado 123 goles y le han anotado 89, entonces esta variable será $123/89$, que nos da un resultado de 1.38. Procedimientos similares los podemos ir aplicando a cada una de las variables que hayamos determinado.

Existen muchos modelos que podemos usar. Una de las áreas con grandes posibilidades se llama *Machine Learning*, o aprendizaje computacional. Dentro de esta área uno de los algoritmos más populares es la red neuronal artificial o NN. Una NN simula burdamente el comportamiento del cerebro, de esta manera tenemos neuronas y conexiones entre ellas. Observe la Figura 1, la cual denota una red neuronal con 2 entradas y una neurona.

Esta figura muestra a 2 equipos, Brasil e Islandia, con el "coeficiente" de goleo que se habría calculado. Las variables que están representadas como W_1 y W_2 , se les conoce como pesos y es donde el modelo aprenderá a reconocer los patrones. Después, la neurona la representamos con la letra griega sigma. Pero, ¿cómo funciona este modelo? Bien, debemos primero "entrenarlo" (qué mejor nombre bajo este contexto). Debemos obtener datos de muchos encuentros de fútbol y formar un conjunto de entrenamiento. Tome en cuenta este ejemplo, la neurona lo que hace con estos 2 equipos es la operación que se muestra en la figura (una sumatoria de 2 multiplicaciones). Nosotros debemos indicar a la neurona que esperar (recuerde que son datos históricos y ya conocemos los resultados). De esta manera si la neurona no obtiene el resultado que le estamos diciendo, los pesos W_1 y W_2 , se modificarán hasta que el resultado sea el esperado. Esto lo hacemos para todos los encuentros de los que dispongamos, al final nos quedaremos con el modelo que tenga menor error. Ahora ya podemos darle los datos de los partidos que se jugarán en el futuro y como la red neuronal ya está entrenada nos dirá quien espera que gane.

Un par de consideraciones, en este modelo que hicimos se toma en cuenta una sola variable, i.e. coeficiente de goleo, pero este modelo se puede generalizar para tomar en cuenta todas las variables que se necesiten. Y otra cosa, la salida de la neurona será un valor entre 0 y 1, es decir, una probabilidad, la cual puede ser usada para ver la certeza de la predicción.

¿A alguien le queda duda de la aplicación de las matemáticas?

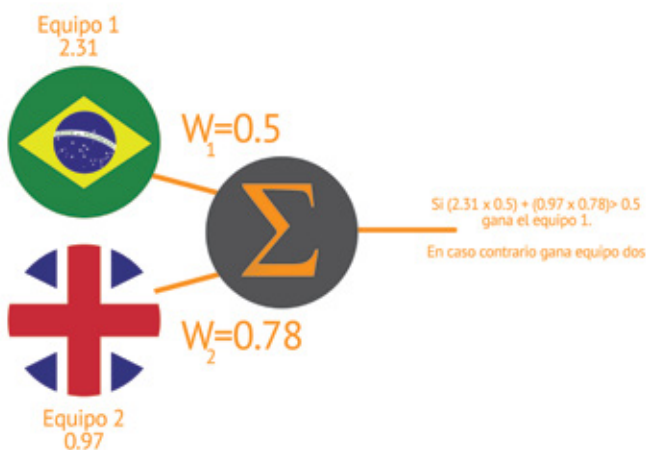


Figura 1. Red neuronal artificial con una neurona y 2 entradas

