

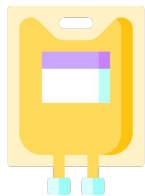
Final project : Predicting postprandial blood glucose level

Mauricio A. De León Cárdenas 505597

Juan M. Álvarez Sánchez 511385

Viviana Vázquez Gómez Martínez 509271

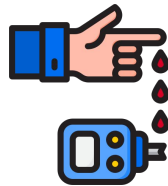
Orlando X. Torres Guerra 513341



Introduccion

- La diabetes es una enfermedad que ocurre cuando la glucosa en la sangre de una persona es muy alta. Aproximadamente, en los Estados Unidos hasta el 2015, 30.3 millones de personas tenían diabetes, pero cerca de 7.2 millones no han sido diagnosticadas, esto debido a que es difícil detectarla por el parecido con otras enfermedades.
- México es el segundo país en el ranking de diabetes a nivel Norteamérica, después de Estados Unidos y sexto a nivel global, con 12.8 millones de personas con esta enfermedad (IDF 2019).
- No identificar la diabetes a tiempo puede llevar a tener serios problemas de salud, así como incrementar la probabilidad de ceguera, falla de riñón o incluso la muerte.
- Existen 2 principales tipos de diabetes, Diabetes Tipo 1 y Diabetes Tipo 2

Diabetes Tipo 1



- La diabetes tipo 1 ocurre cuando el cuerpo no produce suficiente cantidad de la hormona que permite a las células del cuerpo absorber y utilizar la glucosa. Esta hormona es llamada insulina.
- Mientras que una persona puede prevenir la diabetes tipo 2 al tener una dieta baja en azúcares y teniendo un estilo de vida activo, la diabetes tipo 1 no es posible prevenirla.
- Una persona con diabetes tipo 1 va a tener que inyectarse insulina por el resto de su vida. El no hacerlo puede resultar en un aumento descontrolado del nivel de azúcar en la sangre, llevándolo a ser peligroso y traer graves complicaciones.

Objetivo



- El objetivo de este trabajo es predecir el pico más alto de nivel de glucosa en la sangre (medido en mg / dl) después de 2 horas de haber comido, dando como datos:
 - Niveles de glucosa antes de comer
 - Carbohidratos a consumir
 - Unidades de insulina
 - Ejercicio (si hizo o no)
 - Hora de la comida
- Esto puede ayudar al tratamiento de diabetes para alertar al paciente si su dosis de insulina estimada es adecuada al nivel que tendrá y esta lo mantendrá en un nivel de glucosa de 180 mg/dl (que es lo ideal) o su el nivel que el paciente tenga como meta.

Descripción de variables



- **Glucose:** Define el nivel de glucosa en la sangre del usuario justo antes de comer (medido en mg / dl).
- **Carbs:** Cantidad de carbohidratos consumidos (medido en gramos)
- **Units of insulin:** Cantidad de insulina que se inyecta previo a la comida (medido con una jeringa BD)
- **Exercise:** Indicador booleano que indica si el usuario hizo ejercicio antes de comer
- **Time:** Hora en la que fue ingerida la comida (medido en formato de 24 horas)
- **Dato de salida:** Nivel de glucosa en sangre 2 horas después de la comida ingerida. (medido en mg / dl).

Estas variables fueron seleccionadas siguiendo la medición de glucosa preprandial (antes de una comida) y postprandial (después de comer), lo que permite a los pacientes determinar la glucosa esperada 2 horas después de ingerir alimentos. De acuerdo con teoría médica, el pico más alto de nivel de glucosa se ve reflejado aproximadamente 2 horas después de la comida, pero puede variar de paciente en paciente.

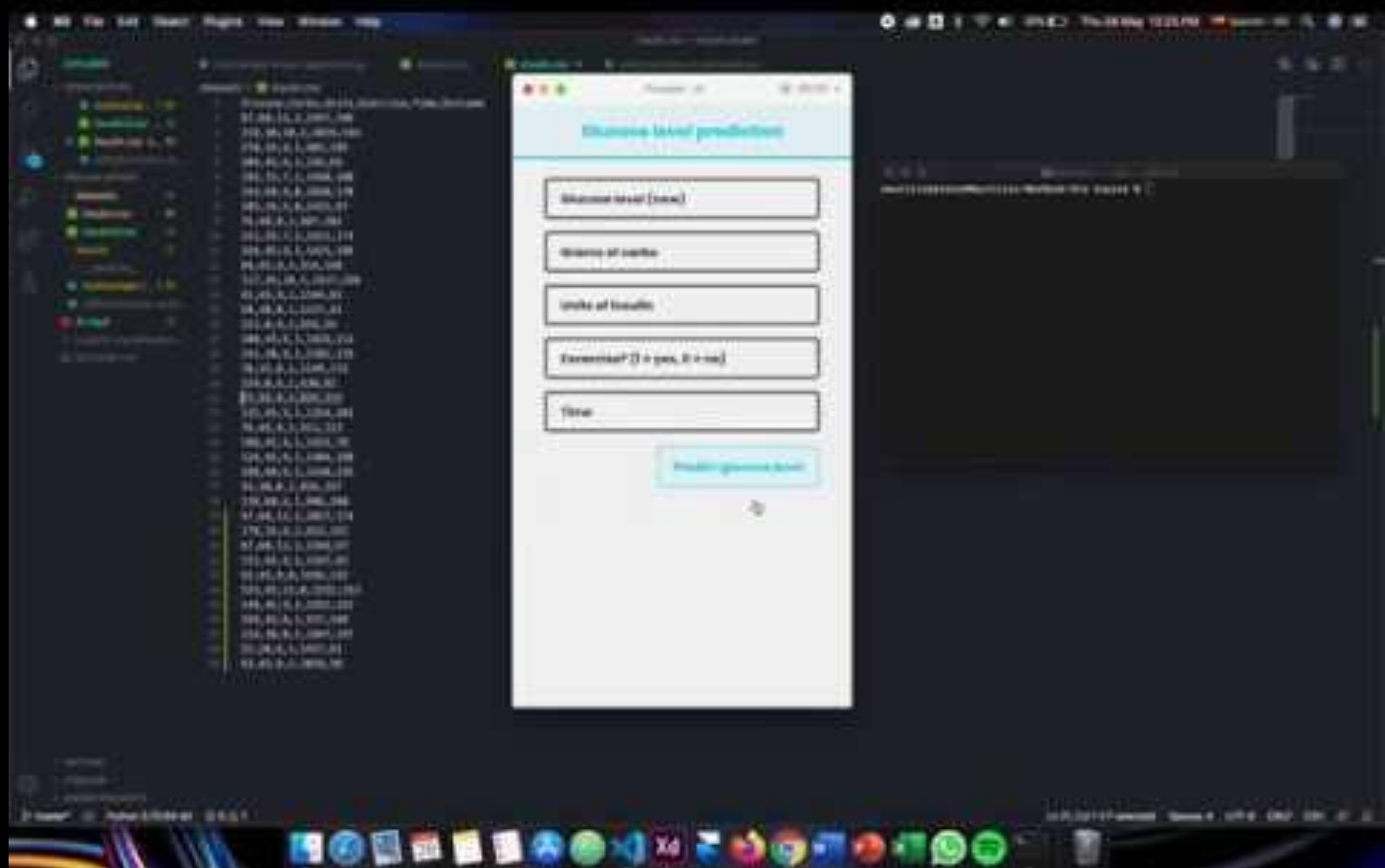
Proceso normal de una persona con diabetes

1. Revisar nivel de glucosa antes de la comida.
2. Hace estimación de los carbohidratos que va a consumir.
3. Basado en los dos datos anteriores, estima cuántas unidades de insulina tiene que inyectarse.
4. A las 2 horas de haber comido revisa su nivel de glucosa y dependiendo el nivel obtenido toma una decisión de si es necesario inyectarse más unidades(en caso de estar por encima de los 180 mg/dl) o consumir algo de azúcar (en caso de estar por debajo de 70mg/dl)



Proceso con ayuda del algoritmo de MLR

1. Revisar nivel de glucosa antes de la comida
2. Hace estimación de los carbohidratos que va a consumir.
3. Registrar estos datos en el algoritmo, así como si hizo ejercicio o no y la hora de la comida.
4. El algoritmo un nivel de glucosa predicho a 2 horas, en el que el usuario podrá ver si es necesario inyectarse más insulina o menos en ese momento.



Desarrollo

Después de identificar los objetivos de este proyecto se decidió observar el dataset, sus features y los valores obtenidos para cada una de estas para decidir si era necesario realizar escalamiento, feature design, or selección de las features.

Al ver los outcomes esperados se optó por utilizar un algoritmo de regresión lineal multivariable el cual utiliza el método del gradiente descendiente que se muestra a continuación:

$$\mathbf{w}_k = \mathbf{w}_{k-1} - \alpha \nabla_{\mathbf{w}} J(\mathbf{w}) = \mathbf{w}_{jk} = \mathbf{w}_{jk-1} - \alpha \frac{1}{N} \sum_{i=1}^N [f_{\mathbf{w}}(\mathbf{x}^{(i)}) - y^{(i)}] \mathbf{x}_j^{(i)}$$

Ya que se obtiene el valor de los pesos de beta (\mathbf{w}) es posible predecir los valores de glucosa esperados utilizando la siguiente fórmula vectorial.

$$f_{\mathbf{w}}(\mathbf{x}) = \mathbf{w}^T \mathbf{x}$$

Resultados

```
-----  
W parameters:  
-----
```

```
w0: [146.81072669]  
w1: [22.08200009]  
w2: [39.41951859]  
w3: [-14.00831021]  
w4: [7.44722397]  
w5: [17.5328232]
```

```
-----  
Predicted glucose outputs  
-----
```

```
[[111.77016708 145.53419839 159.26776769 189.23231259 190.12971012  
 78.1998524 120.10615526]]
```

Testing sample	Actual output	Predicted output
1	142	111.770
2	253	145.534
3	152	159.268
4	168	189.232
5	197	190.130
6	61	78.200
7	58	120.107

Testing Results

Conclusión - Mauricio

A pesar de que los resultados obtenidos no fueron tan buenos como se esperaba, creo realmente que es por la cantidad total de datos que se utilizaron para entrenar el algoritmo, a mayor cantidad de datos mayor cantidad de iteraciones del gradiente descendente y por lo tanto mejor será la estimación de los parámetros w .

Como nota, si este algoritmo llega a ser implementado dentro de una aplicación que les permite a sus usuarios o pacientes determinar su nivel de comida después de cada comida, el algoritmo tendría que entrenarse desde 0 con cada paciente, esto porque aunque la teoría médica indique que el punto alto sucede dos horas después de la comida, esto no siempre es verdad y varía de persona a persona y de su tipo de diabetes. Creo que hacer un uso práctico de este algoritmo podría llegar a ayudar a muchísimas personas con diabetes que todavía no tienen idea (o muy poca) de cómo medir sus niveles de glucos y qué esperar si comen cierto tipo de comidas (esto incluso podría volverse una nueva feature en siguientes mejoras del algoritmo).

Conclusión - Juan Manuel

La implementación del modelo fue buena para los datos disponibles. Se demostró una predicción precisa para los valores entre 60 y 200 mg/dl, la mayoría de las muestras del dataset.

Sin embargo, para una aplicación real, se requiere de mucha disciplina por parte del paciente para recolectar todos los datos necesarios para obtener una buena predicción. Conforme haya más datos con valores altos o bajos de glucosa, las predicciones para este tipo de valores atípicos debería mejorar.

Cabe mencionar que esta implementación tiene un error humano que es la estimación de carbohidratos que se consumen, ya que no es posible conocer el valor exacto de carbohidratos consumidos a menos que el alimento sea completamente empaquetado y etiquetado.

Este algoritmo es útil para personas recién diagnosticadas por que llevan un control más estricto y sí pueden generar todos los datos requeridos para tener predicciones que les ayuden en su proceso de adaptación a su vida con diabetes tipo 1.

Conclusión - Viviana

Como se mencionó, México es el segundo país en norteamérica con mayor número de personas con diabetes, esto se ve reflejado en el estado de vida de los mexicanos. Los resultados obtenidos fueron buenos para la falta de datos que teníamos, puesto que no existe un dataset con los requerimientos que necesitábamos, así que nuestro compañero Juan Manuel nos hizo el favor de recolectar sus datos para este propósito.

Siento que esto es el comienzo de una investigación mayor y una nueva implementación en el área, el algoritmo puede seguir siendo entrenado y colocarse mayor cantidad de datos para que tenga mayor exactitud y pueda ser implementado en una aplicación.

Sin embargo, nos hace falta saber más sobre el tema de la diabetes ya que es una enfermedad peligrosa y 7 de cada 10 mexicanos cuentan con diabetes.

Conclusión - Orlando

Creo que la regresión multilíneal es un algoritmo muy útil para hacer predicciones relacionadas a temas como la diabetes. En este caso, nosotros buscamos predecir el nivel de glucosa 2 horas después de que ingiriera una comida, considerando las variables antes mencionadas. Este algoritmo le puede ser de utilidad a una persona con diabetes ya que le ayudará a medir de mejor manera la cantidad de insulina que se inyectará, llevando así un mejor control de su nivel de azúcar en la sangre.

Considero que el algoritmo fue bien implementado ya que obtuvimos buenos resultados, sin embargo, una de las limitaciones que tuvimos fue que no conseguimos acceso a alguna base de datos que nos fuera de utilidad para evaluar las variables que utilizamos por lo que tuvimos que tomar datos nosotros mismos (nuestro compañero Juan Manuel) y debido al poco tiempo y que solo se pueden recuperar 3 datos diarios, no fueron suficientes datos para que el modelo fuera más preciso.

Referencias

- Nall, R. (2020). *An overview of diabetes types and treatments*. Retrieved on april 19th 2020 from: <https://www.medicalnewstoday.com/articles/323627>
- World Health Organization. (2018). *Diabetes*. Retrieved on april 26 2020 from: <https://www.who.int/news-room/fact-sheets/detail/diabetes>
- National Institute of Diabetes and Digestive and Kidney Diseases. (2016). *What is Diabetes?*. Retrieved on april 26 2020 from: <https://www.niddk.nih.gov/health-information/diabetes/overview/what-is-diabetes>
- Miller, K. (2017). *Nearly a Quarter of People with Diabetes Don't Know They Have It*. Retrieved on april 26 2020 from: <https://www.self.com/story/undiagnosed-diabetes>
- International Diabetes Federation, (2019). *IDF Diabetes Atlas. 9th edition*. [Internet] Available at: <https://www.diabetesatlas.org/en/resources/>
- Swaminathan, S. (2018). *Linear regression - Detailed view*. Retrieved on april 26 2020 from: <https://towardsdatascience.com/linear-regression-detailed-view-ea73175f6e86>
- Kenton, W. (2019). *Multiple Linear Regression - MLR Definition*. Retrieved on april 26 2020 from: <https://www.investopedia.com/terms/m/mlr.asp>
- Medical News Today (2018). *Type 1 diabetes*. Retrieved on april 26 2020 from: <https://www.medicalnewstoday.com/articles/323729>

Código de Honor

Damos nuestra palabra de que hemos realizado este proyecto con integridad académica.