Notas de topología y sus aplicaciones

César Pérez
Universidad De Las Americas Puebla
cesar.perezar@udlap.mx

Índice

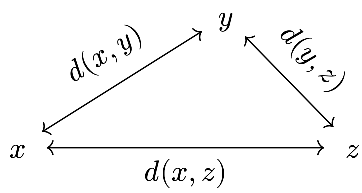
1. Espacios métricos

Una gran mayoría de los ejemplos que usamos para los conceptos del curso vienen del análisis topológico de datos, una rama en crecimiento de la topología aplicada. Cuando analizamos datos nos interesa medir distancias en nuestro muestreo (por ejemplo, el error de un modelo de regresión lineal multiple se mide tomando las distancias de cada medición real con la predicción), y para eso necesitamos una buena noción de lo que significa medir, en concreto una buena noción de *distancia* (Spoiler: **métricas**). Además nos interesa que esta buena noción de distancia pueda relacionar cosas parecidas como cercanas.

1.1. Métricas

Definición 1.1. (Métrica) Sea X un conjunto. una **métrica** es una función $d : X \times X \to \mathbb{R}$, tal que:

1. $d(x, y) \geq 0$
2. $d(x, y) \Leftrightarrow x = y$
3. $d(x, y) = d(y, x)$
4. **Desigualdad del triángulo:** $d(x, z) \leq d(x, y) + d(y, z)$



Ejemplo. Tomemos como «métrica» el tiempo que toma llegar de un punto a otro, sabemos que el tiempo transcurrido siempre es positivo, por lo tanto 1. se cumple. Si nos toma 0 unidades de tiempo llegar a un punto significa que estamos parados justamente en ese punto, entonces 2. se cumple. Sin embargo, el tiempo que nos toma ir de un punto x a un punto y no siempre es igual que el tiempo que nos toma ir de y a x, por lo tanto este ejemplo **no** es una métrica.

Nota. A esta pseudo-métrica se le conoce como la **métrica del alpinista**, ya que para un alpinista toma más tiempo llegar a la sima de una montaña que a la cima.

Proposición 1.2. (Distancia en los reales) Sea $d : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ la función dada por $d(x, y) = |x - y|$. Entonces d define una métrica sobre $\mathbb{R}$.

Demostración. Sea $d(x, y) = |x - y|$ para $x, y \in \mathbb{R}$

1. Por definición de valor absoluto $|x - y| = d(x, y) \geq 0$.
2. Si $x = y$ entonces $x - y = 0$ y por lo tanto $|x - y| = d(x, y) = |0| = 0$. Además si $|x - y| = d(x, y) = 0$ entonces por definición de valor absoluto $x - y = 0$ lo que implica $x = y$.
3. Por propiedades de valor absoluto $d(x, y) = |x - y| = |y - x| = d(y, x)$.
4. Desigualdad del triángulo: $d(x, z) = |x - z| = |(x - y) + (y - z)|$, por desigualdad triángular del valor absoluto en $\mathbb{R}$, para cualesquiera números $a, b \in \mathbb{R}$ se cumple que $|a + b| \leq |a| + |b|$. Sean $a = (x - y), b = (y - z)$ obtenemos

$$d(x, z) = |(x - y) + (y - z)| \leq |x - y| + |y - z|$$

En terminos de la función: $d(x, z) \leq d(x, y) + d(y, z)$

Se concluye que d define una métrica sobre $\mathbb{R}$. □

Nota. Para 4. utilizamos una desigualdad auxiliar, en muchas de las demostraciones de funciones como métricas para demostrar la desigualdad del triángulo vamos a utilizar este tipo de desigualdades.

Proposición 1.3. (Métrica de Manhattan) Sean $\vec{x} = (x_1, ..., x_n), \vec{y} = (y_1, ..., y_n) \in \mathbb{R}^n$ y $d : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ la función dada por:

$$d(\vec{x}, \vec{y}) = \sum_{i=1}^{n} |x_i - y_i|$$

Entonces d define una métrica sobre $\mathbb{R}^n$

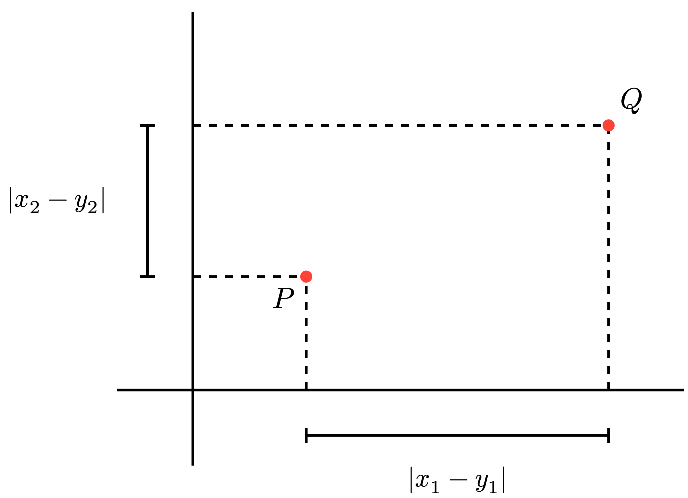
Demostración. Sean $\vec{x} = (x_1, ..., x_n), \vec{y} = (y_1, ..., y_n), \vec{z} = (z_1, ..., z_n) \in \mathbb{R}^n, \alpha \in \{1, ..., n\}$ y

$$d(\vec{x}, \vec{y}) = \sum_{i=1}^{n} |x_i - y_i|$$

1. Cada uno de los sumandos $|x_\alpha - y_\alpha| \geq 0$ entonces
$$d(\vec{x}, \vec{y}) = \sum_{i=1}^{n} |x_i - y_i| \geq 0$$
2. Si $\vec{x} = \vec{y}$ entonces $x_\alpha = y_\alpha$ por lo que $x_\alpha - y_\alpha = 0$, por lo tanto
$$d(\vec{x}, \vec{y}) = \sum_{i=1}^{n} |x_i - y_i| = \sum_{i=1}^{n} |0| = 0$$
Además, tenemos que si $d(\vec{x}, \vec{y}) = 0$ entonces cada uno de los sumandos $|x_\alpha - y_\alpha| = 0$, y $x_\alpha = y_\alpha$ por lo tanto $\vec{x} = \vec{y}$
3. Como $|x_\alpha - y_\alpha| = |y_\alpha - x_\alpha|$ entonces
$$d(\vec{x}, \vec{y}) = \sum_{i=1}^{n} |x_i - y_i| = \sum_{i=1}^{n} |y_i - x_i| = d(\vec{y}, \vec{x})$$
4. Por Proposición 1.2 tenemos que $|x_\alpha - z_\alpha| \leq |x_\alpha - y_\alpha| + |y_\alpha - z_\alpha|$ entonces por axiomas de orden en $\mathbb{R}$ tenemos que
$$\left(d(x, z) = \sum_{i=1}^{n} |x_i - z_i|\right) \leq \left(\sum_{i=1}^{n} |x_i - y_i| + \sum_{i=1}^{n} |y_i - z_i| = d(x, y) + d(y, z)\right)$$

Se concluye que d define una métrica sobre $\mathbb{R}^n$ □

Nota. Se le conoce como métrica de Manhattan ya que en $\mathbb{R}^2$ al visualizar la métrica se «parece» a las calles de Nueva York, en el sentido que es de *izquierda a derecha* y *arriba a abajo* y no en diagonal:



Proposición 1.4. (Distancia euclidiana) Sean $\vec{x} = (x_1, ..., x_n), \vec{y} = (y_1, ..., y_n) \in \mathbb{R}^n$ y $d : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ la función dada por:

$$d(\vec{x}, \vec{y}) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

Entonces d define una métrica sobre $\mathbb{R}^n$

Borrador de la demostración. Sean $\vec{x} = (x_1, ..., x_n) \in \mathbb{R}^n, \vec{y} = (y_1, ..., y_n) \in \mathbb{R}^n, \alpha \in \{1, ..., n\}$ y

$$d(\vec{x}, \vec{y}) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

1. Cada uno de los sumandos $(x_\alpha - y_\alpha)^2 \geq 0$ entonces

$$d(\vec{x}, \vec{y}) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \geq 0$$

2. Si $\vec{x} = \vec{y}$ entonces $x_\alpha = y_\alpha$ por lo que $x_\alpha - y_\alpha = 0$, por lo tanto

$$d(\vec{x}, \vec{y}) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} = \sqrt{\sum_{i=1}^{n} (0)^2} = 0$$

   Además, tenemos que si $d(\vec{x}, \vec{y}) = 0$, cada uno de los sumandos $(x_\alpha - y_\alpha)^2 = 0$ y $x_\alpha = y_\alpha$ por lo tanto $\vec{x} = \vec{y}$
3. Como $(x_\alpha - y_\alpha)^2 = (y_\alpha - x_\alpha)^2$ entonces

$$d(\vec{x}, \vec{y}) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} = \sqrt{\sum_{i=1}^{n} (y_i - x_i)^2} = d(\vec{y}, \vec{x})$$

4. To be continued (Spoiler: **Desigualdad de Cauchy-Schwarz**)

□*

Proposición 1.5. (Métrica p-esima) Sean $\vec{x} = (x_1, ..., x_n), \vec{y} = (y_1, ..., y_n) \in \mathbb{R}^n$ y $d : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ la función dada por

$$d(\vec{x}, \vec{y}) = \left(\sum_{i=1}^{n} |x_i - y_i|^p\right)^{\frac{1}{p}}$$

Entonces d define una métrica sobre $\mathbb{R}^n$

Para las 3 primeras propiedades de métrica la demostracíon de Proposición 1.5 es muy similar a la de Proposición 1.3, sin embargo, como sucedio con la Proposición 1.4, necesitamos una desigualdad auxiliar un poco más compleja para demostrar la desigualdad del triángulo, entonces, la demostración se deja en pausa para poder continuar con los contenidos del capitulo.

Nota. La métrica de Manhattan y la distancia euclidiana son casos particulares de la métrica p-esima, en donde respectivamente $p = 1$ y $p = 2$, además la métrica p-esima nos define una infinidad de medir distancias en $\mathbb{R}^n$.

Notación. Para diferenciar cada una de las métricas que hemos definido, se usa la notación d_p es decir, para la métrica de Manhattan usamos d_1 para la distancia euclidiana usamos d_2 y para la p-esima d_p

Proposición 1.6. (Métrica de Chebyshev) Sean $\vec{x} = (x_1, ..., x_n), \vec{y} = (y_1, ..., y_n) \in \mathbb{R}^n$ y $d_\infty : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ la función dada por:

$$d_\infty(\vec{x}, \vec{y}) = \max_{1 \leq i \leq n} |x_i - y_i|$$

Entonces d_∞ define una métrica sobre $\mathbb{R}^n$

Demostración. Sean $\vec{x} = (x_1, ..., x_n), \vec{y} = (y_1, ..., y_n), \vec{z} = (z_1, ..., z_n) \in \mathbb{R}^n, \alpha \in \{1, ..., n\}$ y

$$d_\infty(\vec{x}, \vec{y}) = \max_{1 \leq i \leq n} |x_i - y_i|$$

1. Cada uno de los candidatos a máximo $|x_\alpha - y_\alpha| \geq 0$ entonces

$$d_\infty(\vec{x}, \vec{y}) = \max_{1 \leq i \leq n} |x_i - y_i| \geq 0$$

2. Si $\vec{x} = \vec{y}$ entonces $x_\alpha = y_\alpha$ por lo que $x_\alpha - y_\alpha = 0$, por lo tanto

$$d_\infty(\vec{x}, \vec{y}) = \max_{1 \leq i \leq n} |x_i - y_i| = \max_{1 \leq i \leq n} |0| = 0$$

   Además si $d_\infty(\vec{x}, \vec{y}) = 0$ entonces cada uno de los candidatos a máximo $|x_\alpha - y_\alpha| = 0$, y $x_\alpha = y_\alpha$ por lo tanto $\vec{x} = \vec{y}$

3. Por propiedades de valor absoluto, cada uno de los candidatos a máximo $(x_\alpha - y_\alpha) = (y_\alpha - x_\alpha)$ entonces

$$d_\infty(\vec{x}, \vec{y}) = \max_{1 \leq i \leq n} |x_i - y_i| = \max_{1 \leq i \leq n} |y_i - x_i| = d_\infty(\vec{y}, \vec{x})$$

4. Sea $d_\infty(\vec{x}, \vec{z}) = |x_\alpha - z_\alpha|$, por Proposición 1.2 $|x_\alpha - z_\alpha| \leq |x_\alpha - y_\alpha| + |y_\alpha - z_\alpha|$ además, si $d_\infty(\vec{x}, \vec{y}) = \max_{1 \leq i \leq n} |x_i - y_i| \neq |x_\alpha - y_\alpha|$ ó $d_\infty(\vec{y}, \vec{z}) = \max_{1 \leq i \leq n} |y_i - z_i| \neq |y_\alpha - z_\alpha|$ entonces

$$d_\infty(\vec{x}, \vec{z}) = |x_\alpha - z_\alpha| \leq |x_\alpha - y_\alpha| + |y_\alpha - z_\alpha| \leq d_\infty(\vec{x}, \vec{y}) + d_\infty(\vec{y}, \vec{z})$$

   Por transitividad $d_\infty(\vec{x}, \vec{z}) \leq d_\infty(\vec{x}, \vec{y}) + d_\infty(\vec{y}, \vec{z})$

Se concluye que d_∞ define una métrica sobre $\mathbb{R}^n$ □

Nota. Todos los ejemplos de métricas que hemos definido son para datos númericos, sin embargo, ¿Cómo medimos distancias entre datos categóricos?

1.2. Métricas para datos categóricos

Definición 1.7. (Dato categórico) Un dato categórico es una variable que toma valores dentro de un conjunto finito y predefinido de etiquetas o clases. Representan cualidades o características, en lugar de cantidades numéricas continuas.

Al no tener cantidades númericas involucradas nuestras métricas definidas anteriormente no pueden medir *directamente* distancias entre datos categóricos. Pero si podemos comparar su **igualdad**.

Proposición 1.8. (Métrica discreta) Sea X un conjunto de datos categóricos, $x, y \in X$ y $d_{\text{disc}} : X \times X \to \mathbb{R}$ la función dada por:

$$d_{\text{disc}}(x, y) = \begin{cases} 0 \text{ si } x = y \\ 1 \text{ si } x \neq y \end{cases}$$

Entonces d define una métrica sobre X

Demostración. Sea X un conjunto de datos categóricos, $x, y, z \in X$ y

$$d_{\text{disc}}(x, y) = \begin{cases} 0 \text{ si } x = y \\ 1 \text{ si } x \neq y \end{cases}$$

1. Como $\text{Im}(d_{\text{disc}}) = \{0, 1\}$ entonces $d_{\text{disc}}(x, y) \geq 0$
2. Por definición de $d_{\text{disc}} : d_{\text{disc}}(x, y) = 0 \Leftrightarrow x = y$
3. Como $x = y \Leftrightarrow y = x$ entonces $d_{\text{disc}}(x, y) = d_{\text{disc}}(y, x)$
4. Si $x = z$ entonces $0 = d_{\text{disc}}(x, z) \leq d_{\text{disc}}(x, y) + d_{\text{disc}}(y, z)$, además si $x \neq z$ y tomamos $x = y$ y $y = z$ por transitividad $x = z(!)$ entonces $1 = d_{\text{disc}}(x, z) \leq d_{\text{disc}}(x, y) + d_{\text{disc}}(y, z)$

Se concluye que d_{disc} define una métrica sobre X □

Ejemplo. Para datos categóricos del tipo «negro ó blanco» esta métrica funciona muy bien, sin embargo, ¿Cómo medimos distancias para cosas como colores? Por ejemplo ¿Hay alguna forma de medir d(verde, azul), mantiendo además la relación parentesco-cercanía? Pues conocemos formas de medir distancias en $\mathbb{R}^3$, entonces nos convendría tener una asociación de la forma f : colores $\to \mathbb{R}^3$. Dicha asociación ya existe y se denota rgb, $x = (r, g, b)$ por la cantidad de rojo, verde y azul de cada color, algo importante es que mantiene los colores parecidos «cerca».

En este caso el problema se resolvió aplicando un proceso llamado «vectorización», tomamos un color (dato cualitativo) y lo vectorizamos con valores númericos basandonos en la cantidad de colores primarios que contienen. En general vectorizar es muy útil ya que nos permite utilizar las métricas que ya definimos sobre $\mathbb{R}^n$, sin embargo hay que recordar que es importante mantener la relación parentesco-cercanía.

Ejemplo. ¿Cómo medimos distancias o más bien «similitud» en dos carritos de Amazon? Podemos vectorizar basandonos en las compras, imaginemos que Amazon solo tiene 8 objetos a la venta y si un usuario compra uno asignamos «1», el resto siendo «0», i.e:

$$\boldsymbol{c_1} = (0, 1, 1, 1, 1, 1, 1, 0)$$
$$\boldsymbol{c_2} = (1, 1, 1, 1, 1, 0, 1, 1)$$

Tomamos la cardinalidad de la diferencia geométrica de los dos vectores, es decir

$$d(\boldsymbol{c_1}, \boldsymbol{c_2}) = \#(\boldsymbol{c_1} \triangle \boldsymbol{c_2}) = 3$$

Nota. Podemos aplicar este módelo de contar en cuántas coordenadas dos objetos son diferentes a muchos problemas.

Proposición 1.9. (Distancia de Hamming) Sean Σ un abecedario finito de la forma $\{\ell_1, ..., \ell_m\}$, $n \in \mathbb{N}$, dos palabras finitas $p_1 = (\ell_1^1, ..., \ell_n^1), p_2 = (\ell_1^2, ..., \ell_n^2) \in \Sigma^n$ y $d_H : \Sigma^n \times \Sigma^n \to \mathbb{N}$ la función dada por:

$$d_H(p_1, p_2) = \#\{i \in \{1, ..., n\} : \ell_i^1 \neq \ell_i^2\}$$

Entonces d_H define una métrica sobre Σ^n

Borrador de la demostración. Sean Σ un abecedario finito de la forma $\{\ell_1, ..., \ell_m\}$, $n \in \mathbb{N}$, dos palabras finitas $p_1 = (\ell_1^1, ..., \ell_n^1), p_2 = (\ell_1^2, ..., \ell_n^2) \in \Sigma^n$ y

$$d_H(p_1, p_2) = \#\{i \in \{1, ..., n\} : \ell_i^1 \neq \ell_i^2\}$$

1. La cardinalidad siempre es un número natural $d_H(p_1, p_2) = \#\{i \in \{1, ..., n\} : \ell_i^1 \neq \ell_i^2\} \geq 0$
2. Si $p_1 = p_2$ entonces $\ell_i^1 = \ell_i^2$ para toda $i \in \{1, ..., n\}$ por lo tanto $\#\{i \in \{1, ..., n\} : \ell_i^1 \neq \ell_i^2\} = \#\emptyset = 0$. Además si $d_H(p_1, p_2) = 0$ entonces $\#\{i \in \{1, ..., n\} : \ell_i^1 \neq \ell_i^2\} = 0$ lo que implica que no existe ningún elemento distinto entre las palabras es decir $p_1 = p_2$
3. Por la definición del conjunto contable $\{i \in \{1, ..., n\} : \ell_i^1 \neq \ell_i^2\} = \{i \in \{1, ..., n\} : \ell_i^2 \neq \ell_i^1\}$ tomando cardinalidades: $d_H(p_1, p_2) = d_H(p_2, p_1)$
4. To be continued (Spoiler: **Distancia en grafos**)

$\square^*$

Proposición 1.10. (Distancia en grafos) Sean $G = (V, E)$ un grafo no orientado y finito, $u, v \in V$ y $d_G : V \times V \to \mathbb{N}$ la función dada por:

$$d_G(u, v) = \min\{k : \text{existe un camino de longitud } k \text{ entre } u, v\}$$

Entonces d_G define una métrica sobre los vertices de G

Demostración. Sean $G = (V, E)$ un grafo no orientado y finito, $u, v, w \in V$ y

$$d_G(u, v) = \min\{k : \text{existe un camino de longitud } k \text{ entre } u, v\}$$

1. Como la longitud de cualquier camino es una suma de aristas de tamaño $k \in \mathbb{N}$ el mínimo de un conjunto de números naturales siempre es mayor o igual a 0. Esto es $d_G(u, v) \geq 0$
2. Si $u = v$, el camino trivial de longitud 0 existe, por lo tanto $d_G(u, v) = 0$. Además si $d_G(u, v) = 0$ el cámino mas corto tiene 0 aristas. Un camino sin aristas empieza y termina en el mismo vertice, es decir $u = v$
3. Al ser un grafo no orientado, si existe un camino más corto de u a v dado por la secuencia de vertices $(u, x_1, ..., x_k, v)$ la secuencia inversa $(v, x_k, ..., x_1, u)$ es un camino de v a u de la misma longitud. Por tanto la distancia mínima es igual, es decir $d_G(u, v) = d_G(v, u)$
4. Sean P_{uv} un camino mínimo entre u y v (longitud $d_G(u, v)$) y P_{vw} un camino mínimo entre v y w (longitud $d_G(v, w)$). Al concatenar estos dos caminos, formamos un camino de u a w que pasa por v, cuya longitud total es $d_G(u, v) + d_G(v, w)$, como $d_G(u, w)$ esta definido como la longitud del camino más corto posible entre u y w este debe ser menor o igual que la longitud de cualquier otro camino arbitrario que construyamos (incluyendo el que pasa por v) por lo tanto

$$d_G(u, w) \leq d_G(u, v) + d_G(v, w)$$

.

Se concluye que d_G define una métrica sobre V $\square$

Nota. Usando la distancia de grafos podemos mapear una equivalencia con la distancia de Hamming y construir un grafo que modele el espacio de palabras para demostrar indirectamente la desigualdad del triángulo de d_H

Lemma 1.11. (Equivalencia de métricas) Construimos un grafo $G = (V, E)$ con $V = \Sigma^n$ y conectamos dos palabras $u, v \in V$ si y solo si difieren exactamente en una posición. Es decir

$$(u, v) \in E \Leftrightarrow d_H(u, v) = 1$$

Entonces $d_H(p_1, p_2) = d_G(p_1, p_2)$

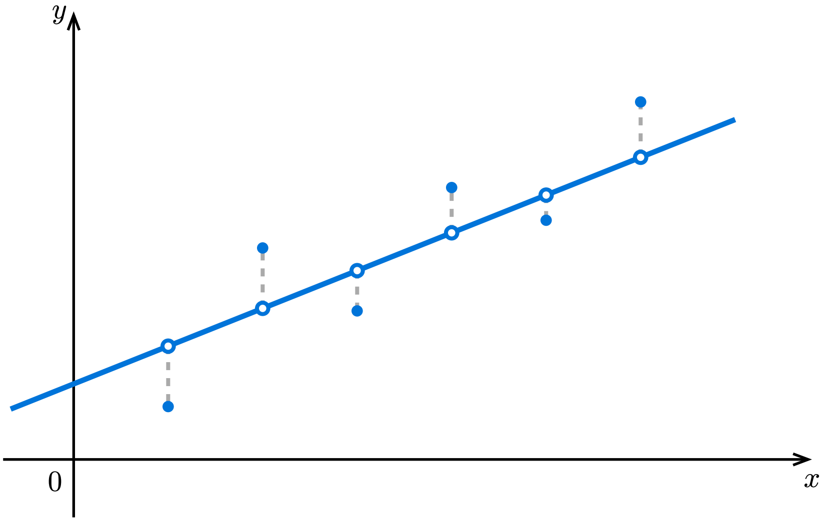
Demostración. (Continuación de d_H) Ya demostramos 1. 2. y 3. Ahora, usando el Lemma 1.11 $d_G = d_H$, como ya demostramos $d_G(p_1, p_3) \leq d_G(p_1, p_2) + d_G(p_2, p_3)$ sustituyendo con d_H:

$$d_H(p_1, p_3) \leq d_H(p_1, p_2) + d_H(p_2, p_3)$$

Se concluye que d_H define una métrica sobre Σ^n □

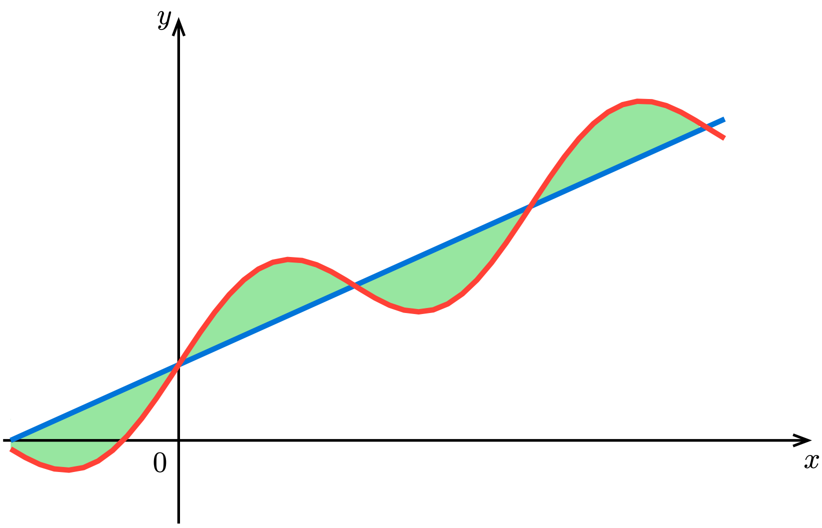
1.3. Distancia entre funciones

Al modelar tenemos que tomar en cuenta la robustez del módelo, es decir, que una pequeña distorción en los datos cause una pequeña distorción en el módelo. Como por lo general el módelo es una función continua podemos tomar lo que varían los resultados reales contra los esperados, es decir, tomar la distancias entre estos.



Suponiendo que nuestro conjunto de datos reales es $R = (r_1, \ldots, r_n)$ y nuestro conjunto de predicciones en esos datos es $Y = (y_1, \ldots, y_n)$ entonces podemos usar la métrica de Manhattan para calcular el error del módelo: $d_1(R, Y) = |r_1 - y_1| + \ldots + |r_n - y_n|$, sin embargo, nos interesa además poder derivar esta expresión para minimizarla, para esto lo que hacemos es elevar al cuadrado cada expresión: $\text{Error} = (r_1 - y_1)^2 + \ldots + (r_n - y_n)^2$

¿Qué pasa con resultados continuos? Entonces podemos representar los resultados como una función; nuestro módelo también es una función, entonces podemos medir distancias como áreas entre curvas. En donde $d_1(e, r) = \int_a^b |e(x) - r(x)| \, \mathrm{d}x$, asumiendo que e es nuestro módelo y r son los datos reales.



En general podemos encontrar la distancia con métricas que ya conocemos, por ejemplo:

$$d_p(e, r) = \left(\int_a^b |e(x) - r(x)|^p \right)^{\frac{1}{p}}$$