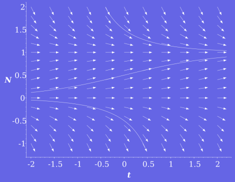
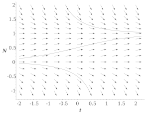
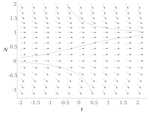
Modelos y Datos

Busquedas locales y globales.



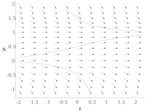
1





Modelos y Datos

- Los datos son un reflejo de la realidad. Afirmar lo contrario no es correcto.
- La forma en que se realiza el proceso de adquisición de datos, impacta la abstracción que se hace de la realidad (el modelo).

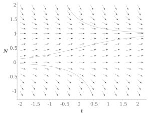


Preparación de datos

Modelo y datos de una epidemia

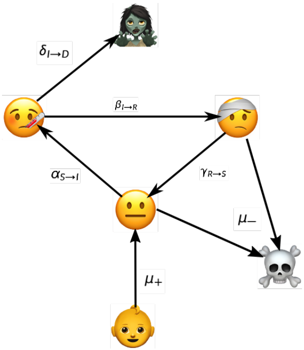
Marco de referencia.

Con propósitos demostrativos, en la siguiente presentación se utilizaran los datos y el modelo para la epidemia del Covid19 tomados para la población de Corea del Sur. Esto debido a su agresiva campaña de pruebas masivas y fiabilidad en su proceso de recopilación de datos.



Modelo a utilizar

Dada la naturaleza de la epidemia, se considera la tasa de natalidad $\mu_+ = 0$. *Por simplicidad consideramos que los susceptibles a Covid19, solamente mueren de una infección de Covid19, por tanto* $\mu_- = 0$. *Asumimos de igual forma que un paciente recuperado alcanza inmunidad, por tanto* $\gamma = 0$.



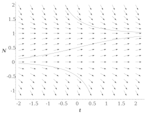
Ecuaciones

$$S_t = N - I_t - D_t - R_t \tag{1}$$

$$\dot{I}_t = a\, I_t S_t - b\, I_t - \delta I_t \tag{2}$$

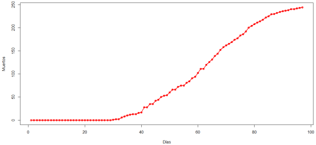
$$\dot{R}_t = b I_t \tag{3}$$

$$\dot{D}_t = \delta I_t \tag{4}$$

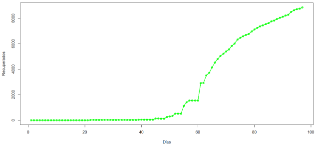


Preparación de datos

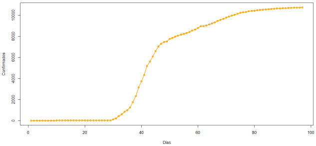
Datos crudos



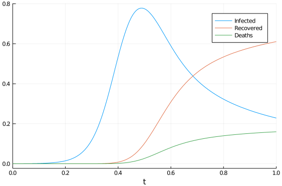
(a) Muertos reportados



(b) Recuperados reportados



(c) Casos Confirmados.

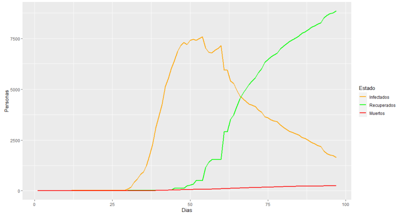


(d) Familia de Soluciones SIRD

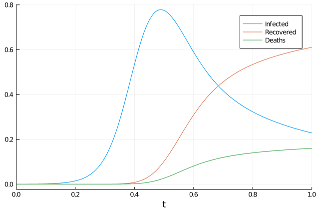
Figure: Datos crudos

Preparación de datos

La forma de los datos importa.

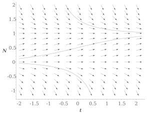


(a) Datos

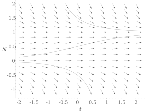


(b) Solución del SIRD.

- Infectados=Confirmados-Recuperados-Muertos



2



Entrenamiento

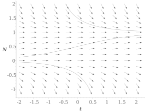
¿Tiene un modelo? ¿Tiene datos? Ajuste el modelo a los datos.

Escoger el modelo es importante

Es importante contar con un **modelo parametrizable**, cuyo resultado sea comparable con los datos por un **criterio de selección de parámetros** conveniente para la situación.

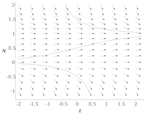
Entrenamiento

Dado un conjunto de observaciones, un modelo y un criterio de comparación, encontrar los parámetros que optimizan el criterio de comparación entre las observaciones y el modelo.



Problemas de optimización

- Al procedimiento de buscar los parámetros correctos de un modelo es también conocido como problemas de inversión o estimación de parámetros.
- La mayoría de problemas de este tipo se encuentran planteados como problemas de optimización. Es decir buscar, el conjunto de parámetros que optimizan el criterio de selección considerando los datos observados.
- Es importante entender las relaciones consideradas entre las variables y los parámetros, así como el alcance de dichas relaciones.



Planteamiento formal del problema

Para nuestro caso en específico, planteamos el problema de la siguiente forma:

Planteamiento del problema.

Encontrar los parámetros p dentro del conjunto Π de parámetros admisibles por el problema. Tal que minimizan la diferencia relativa, normalizada entre el vector de observaciones en función del tiempo, $\mathbf{f} = (f_1, \ldots, f_N)$, y la salida del modelo $\hat{\mathbf{f}}(\cdot; p) = \left(\hat{f}_1(\cdot; p), \ldots, \hat{f}_N(\cdot; p)\right)$. Es notación,

$$\min_{p \in \Pi} \sum_{i=1}^{N} \frac{\left\| f_i(t) - \hat{f}_i(t; p) \right\|_1}{\| f_i(t) \|_1} \tag{5}$$

sujeto a las soluciones del esquema de aproximación:

$$\mathscr{D} \hat{\mathbf{f}}(t; p) = 0, \tag{6}$$

que modela la dinámica de la epidemia.

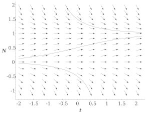
Problemas de busqueda

(a) En una acera

(b) En un parque

(c) En el interior de un rascacielo

Figure: Busqueda



Busqueda local vs Busqueda global

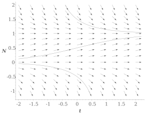
En general, los métodos de busqueda se pueden encontrar en dos clases.

Búsqueda Local

Los métodos de búsqueda local, resuelven el problema buscando puntos estacionarios de la función de costo que representa el criterio de selección. (Informalmente, un punto estacionario es aquél donde la razón de cambio entre puntos cercanos a él es cero).

Búsqueda Global

Por lo general los métodos de búsqueda global, exploran todo el espacio, para encontrar el mejor conjunto de parámetros. La mayoría de estos métodos utilizan heurística para solucionar el problema y son caros computacionalmente.



Búsqueda Global.

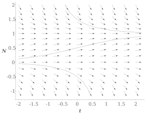
En algunas ocasiones es posible resolver un problema de optimización global de forma cerrada. Es decir, el problema posee una estructura y determinadas condiciones de las cuales es posible obtener una expresión para calcular el óptimo global de forma directa. Por ejemplo, la regresión lineal con la medida de error de los mínimos cuadrados.

Problema.

Dados Y un vector de observaciones y una matriz de variables X. Sea ϵ una variable aleatoria de media cero que representa la diferencia $\epsilon = Y - X\vec{\beta}$. Encontrar los parámetros $\vec{\beta}$ que minimizan la varianza de $\mathrm{Var}[\epsilon] = \mathrm{E}\left[\|Y - X\vec{\beta}\|^2\right]$.

Solución.

La solución del problema esta dictaminada por la ecuación: $\vec{\beta} = \left(X^\intercal X\right)^{-1} X^\intercal Y$



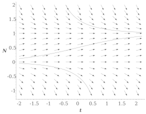
Búsqueda global

¿Podemos modelar nuestro problema con una regresión?

Si tomaramos m observaciones en **condiciones ideales de una dinámica idéntica al modelo** en cada uno de los días, la ecuación del lado derecho cumple.

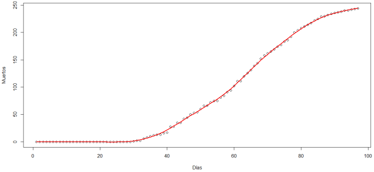
Replanteamiento

$$\begin{pmatrix} \dot{I}_1 \\ \vdots \\ \dot{I}_m \\ \dot{R}_1 \\ \vdots \\ \dot{R}_m \\ \dot{D}_1 \\ \vdots \\ \dot{D}_m \end{pmatrix} = \begin{pmatrix} I_1 S_1 & -I_1 & -I_1 \\ \vdots & \vdots & \vdots \\ I_m S_m & -I_m & -I_m \\ 0 & I_1 & 0 \\ \vdots & \vdots & \vdots \\ 0 & I_m & 0 \\ 0 & 0 & I_1 \\ \vdots & \vdots & \vdots \\ 0 & 0 & I_m \end{pmatrix} \begin{pmatrix} a \\ b \\ \delta \end{pmatrix}$$

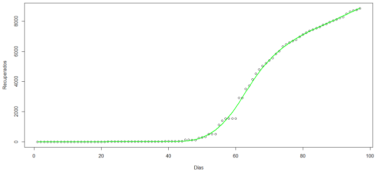


Tratando datos

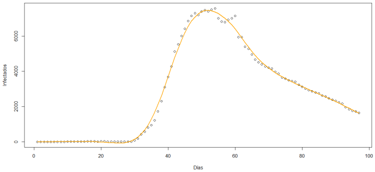
Suavizando las curvas.



(a) Curva de Muertos



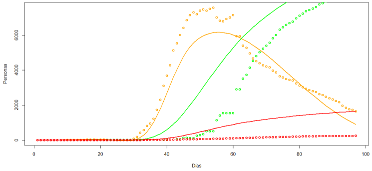
(b) Curva de Recuperados



(c) Curva de Activos

Resultado del modelo

Resultado del modelo

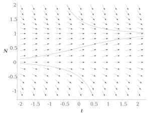


Los valores obtenidos bajo esta estimación:

$$a = 23.2753$$
$$b = 3.63$$
$$c = 0.523072$$

Consideraciones

- Es imposible estimar la población total (personas que tuvieron contacto con los infectados).
- Se escogió un valor que ajustara las gráficas, considerando que los modelos alcanzan puntos de saturación, siendo 10,000 por practicidad dicho número.
- Al usar una regresión, el error a optimizar es independiente de la magnitud. Se compara de igual forma el error en el cálculo de muertos que recuperados.

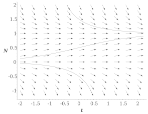


¿Es posible hacerlo mejor?

Reconsiderar el modelo.

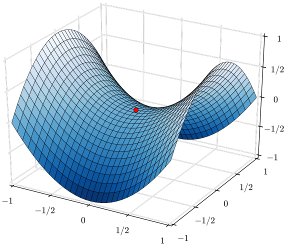
Considerar que las curvas reales contra las curvas del modelo presentan un desfase mayor. La forma de interpretar este resultado es que el modelo considera que los individuos cambian de estado en forma proporcional estado actual. Sin embargo, las personas tardan en sanar pero pueden infectar inmediatamente. De esta forma podemos tanto es posible considerar la siguiente dinámica.

$$
\begin{aligned}
S_t &= N - I_t - D_t - R_t \\
\dot{I}_t &= a\, I_t S_t - b\, I_{t-\tau} - \delta I_{t-\tau} \\
\dot{R}_t &= b I_{t-\tau} \\
\dot{D}_t &= \delta I_{t-\tau}
\end{aligned}
$$

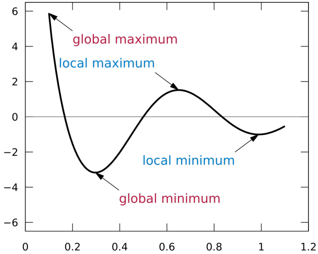


Búsqueda local

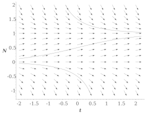
- Dado un punto inicial, buscan la dirección de máximo descenso (o ascenso). Por esta razón, son conocidos como *Exploitatory methods*.
- Pueden llevar a puntos sillas. Puntos que no son ni mínimos ni máximos, pero que son el final del camino de máximo descenso.



(a) Punto silla



(b) Puntos Extremos



Gradientes

El gradiente es la dirección de máximo descenso (o ascenso). La idea general es escoger un punto cercano al punto óptimo y moverse en esa dirección. Una vez en la nueva posición, se calcula nuevamente el gradiente, y nos movemos en esa dirección. Se repite el proceso hasta satisfacer algún criterio de convergencia.

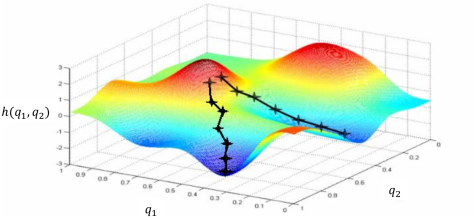
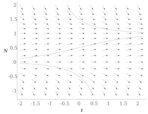


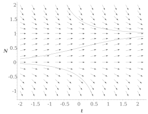
Figure: Ejemplos de búsquedas locales por gradiente



Gradientes de ecuaciones diferenciales

Julia y Diffeqsensitivity

- Julia es un lenguaje de programación con principal enfoque en computación de alto rendimiento.
- Posee librerías para aproximación de ecuaciones diferenciales y paquetes de optimización.
- La librería `Diffeqsensitivity` permite calcular el gradiente del esquema de aproximación de la ecuación diferencial.



Resultado

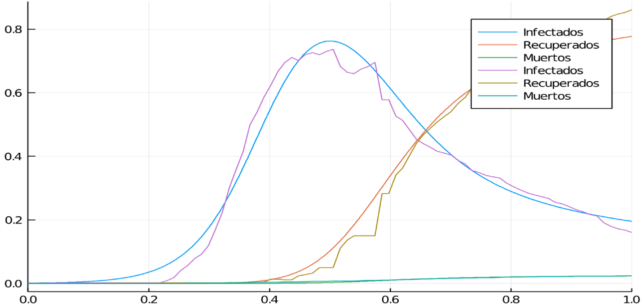
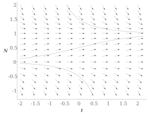


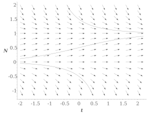
Figure: Modelo Ajustado

$$a \approx 17.703 \quad b \approx 5.91 \quad c \approx 0.179 \quad \tau \approx 12.314 \quad N \approx 10,284 \quad (7)$$



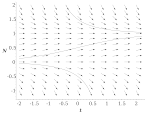
3

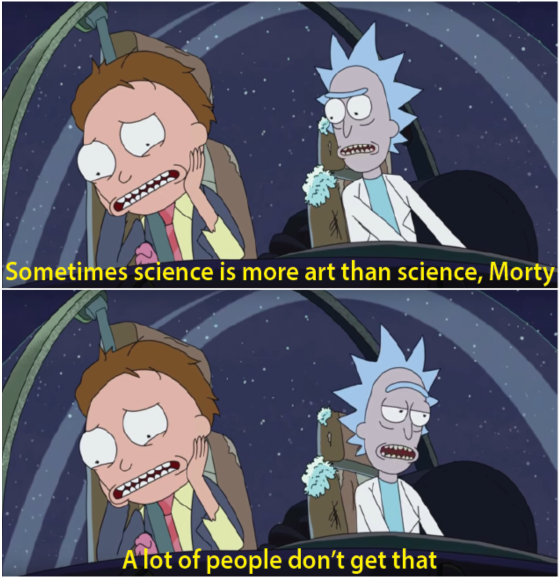
Bibliography



[1] Fabio A.C.C.Chalub Max O.Souzab. *The SIR epidemic model from a PDE point of view*. Elsevier, Mathematical and Computer Modelling, Volume 53, Issues 7–8, April 2011.

[2] Chris Rackauckas.*Basic Parameter Estimation, Reverse-Mode AD, and Inverse Problems. October, 2019*. https://mitmath.github.io/18337/lecture10/estimation$_i$*dentification*




Sometimes science is more art than science, Morty
A lot of people don't get that