UVG UNIVERSIDAD DEL VALLE DE GUATEMALA

Otros Métodos de Regresión

Alan Reyes-Figueroa

Aprendizaje Estadístico

(Aula 30) 22.mayo.2024

Regresión Lineal

¿Qué hacer cuando fallan los supuestos?

- Otras alternativas del modelo de regresión (modelos paramétricos más complejos).
- Transformar los datos (log o transformaciones Box-Cox),
- Regresión lineal no-paramétrica / robusta,
- Otros modelos de regresión:
  - *SVM Regression*,
  - *Random Forests Regression*,
  - Regresión Kernel,
  - Redes neuronales.

Modelos más generales de regresión

La regresión lineal ordinaria (OLS), es un caso especial dentro de una familia de métodos o filosofías de más generales:

- Regresión Lineal Pesada (WLS)
- Regresión no-lineal (optimización Gauss-Newton)
- Regresión lineal generalizada (GLS)
- Regresión con métodos de regularización:
  - Regresión Ridge
  - Regresión LASSO
  - ...
- ...

Regresión Lineal Pesada

La **regresión lineal pesada** o **mínimos cuadrados pesados** (WLS) es una generalización del modelo de regresión lineal, en la que el conocimiento de la varianza de cada observación se incorpora al modelo.

La regresión lineal pesada, ocurre como un caso especial del modelo general de regresión lineal cuando la matriz de covarianza de los residuales ε_i es diagonal

$$\mathbf{W} = \begin{pmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_n^2 \end{pmatrix}.$$

Regresión Lineal Pesada

El residual en el modelo es $\epsilon_i = y_i - \widehat{y}_i = y_i - f(\mathbf{x}_i, \beta)$. Si en el modelo todos los residuales son no-correlacionados y de igual varianza σ^2, entonces los coeficientes óptimos se encuentran mediante el método de máxima verosimilitud $\frac{\partial \mathcal{L}}{\partial \beta} = 0$.

Cuando las mediciones están no-correlacionadas, pero con varianzas distintas, se debe adoptar un método diferente. Aitken mostró que cuando una suma ponderada de residuos cuadrados es minimizada, el mejor estimador lineal insesgado (BLUE) se encuentra cuando cada peso es igual al inverso de la varianza:

$$J = \sum_{i=1}^{n} W_{ii}(y_i - \widehat{y}_i)^2, \quad \text{con } W_{ii} = \frac{1}{\sigma_i^2}.$$

Regresión Lineal Pesada

Así, definimos nuestro funcional de error cuadrático como

$$J = \sum_{i=1}^{n} W_{ii}\,\varepsilon_i = \varepsilon^T \mathbf{W}\varepsilon = ||\mathbf{W}^{1/2}\varepsilon||^2 = ||\mathbf{W}^{1/2}(y_i - \mathbb{X}\beta)||^2.$$

En ese caso, el gradiente está dado por

$$\nabla_\beta J = -2\langle \mathbf{W}^{1/2}\mathbb{X}, \mathbf{W}^{1/2}(\mathbf{y} - \mathbb{X}\beta)\rangle = (\mathbf{W}^{1/2}\mathbb{X})^T\mathbf{W}^{1/2}(\mathbf{y} - \mathbf{W}\mathbb{X}\beta) = \mathbb{X}^T\mathbf{W}(\mathbf{y} - \mathbb{X}\beta) = 0.$$

Obtenemos así las **ecuaciones normales modificadas**

$$\boxed{\mathbb{X}^T\mathbf{W}\mathbb{X}\beta = \mathbb{X}^T\mathbf{W}\mathbf{y}} \quad \Rightarrow \quad \widehat{\beta} = (\mathbb{X}^T\mathbf{W}\mathbb{X})^{-1}\mathbb{X}^T\mathbf{W}\mathbf{y}.$$

Esto equivale a aplicar la regresión lineal ordinaria (OLS) con los datos transformados $\mathbb{X}' = \mathbf{W}^{1/2}\mathbb{X}$ y $\mathbf{y}' = \mathbf{W}^{1/2}\mathbf{y}$.

Regresión Lineal Generalizada

En el **modelo lineal generalizado** (GLM), se asume que la variable dependiente **y** se genera a partir a una distribución en familia exponencial (incluye a la normal, binomial, Poisson, gamma, ...)

Asume que la media μ de la distribución depende de las v.a. independientes X

$$\mathbb{E}(Y \mid X) = g^{-1}(\mathbb{X}\beta),$$

donde g es una función, y que la varianza es una función V de la media

$$Var(Y \mid X) = V(\mu) = V(g^{-1}(\mathbb{X}\beta)),$$

donde $V \sim f$ sigue una distribución f en la familia exponencial.

Componentes del modelo GLM:

- Una distribución f perteneciente a la familia exponencial,
- Un predictor lineal $\eta = \mathbb{X}\beta$,
- Una función de enlace g, tal que $\mathbb{E}(Y \mid X) = \mu = g^{-1}(\eta)$.

Regresión Lineal Robusta

También se les llama métodos de regresión no-paramétrica.

- Estimador de Theil-Sen.
- *Repeated Median Regression*,
- *Iteratively reweighted least squares*,
- Estimador M,
- *Relaxed intersection*,
- RANSAC,
- ...

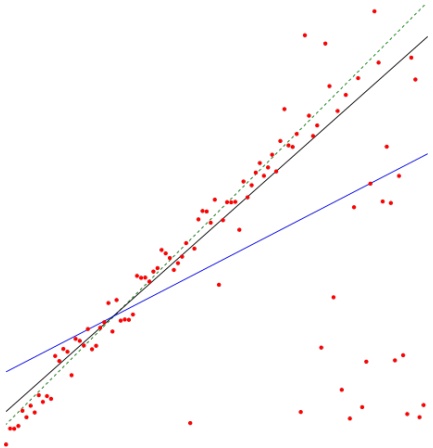
Estimador de Theil-Sen

Definido por Theil (1950). Llamado también el **estimador de Sen**, **método simple de medianas**, o **método robusto de Kendall**.

El estimador de Theil-Sen de un conjunto de puntos bidimensionales (x_i, y_i) es la mediana m de las pendientes determinadas por todos los pares de muestras puntos (con $x_i \neq x_j$). Sen (1968) amplió esta definición para manejar el caso $x_i = x_j$.

Una vez que se ha determinado la pendiente m, se determina el intercepto b como la mediana de los valores $y_i - mx_i$.
Se puede determinar intervalos de confianza para la estimación de la pendiente como el intervalo que contiene el $1 - \alpha$ porcentaje medio de las pendientes.

Estimador de Theil-Sen



Estimador de Theil-Sen (negro), mínimos cuadrados (azul). En verde el *ground-truth*.

Repeated median regression

La **regresión mediana repetida** es un algoritmo de regresión lineal robusto. Este método estima la pendiente m de la recta de regresión $y = mx + b$ para un conjunto de puntos (x_i, y_i) como

$$\widehat{m} = \text{mediana}_i \ \text{mediana}_{j \neq i} \ m_{ij} = \text{mediana}_i \ \text{mediana}_{j \neq i} \ \frac{y_j - y_i}{x_j - x_i},$$

donde $m_{ij} = \frac{y_j - y_i}{x_j - x_i}$ es la pendiente de la recta entre los puntos (x_i, y_i) y (x_j, y_j).

La intersección estimada b con el eje y se define como

$$\widehat{b} = \text{mediana}_i \ \text{mediana}_{j \neq i} \ b_{ij} = \text{mediana}_i \ \text{mediana}_{j \neq i} \ \frac{x_j y_i - x_i y_j}{x_j - x_i},$$

donde $b_{ij} = \frac{x_j y_i - x_i y_j}{x_j - x_i}$ es el intercepto de la recta entre (x_i, y_i) y (x_j, y_j).

RANSAC

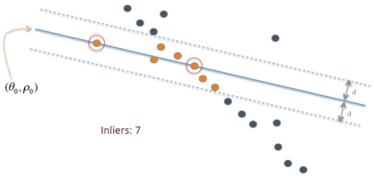
Random sample consensus (RANSAC) es un método iterativo para calcular los parámetros de un modelo de un conjunto de datos que contiene valores atípicos. Es un algoritmo no determinista en el sentido de que produce un resultado razonable sólo con una cierta probabilidad, mayor a medida que se permiten más iteraciones. Fue publicado por Fischler y Bolles (1981).

Algoritmo (RANSAC)

Repetir:

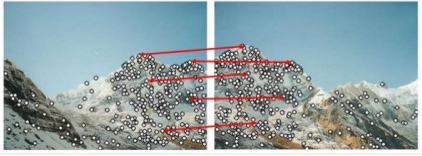
1. Seleccionar una muestra aleatoria S de los datos originales (*inliers* hipotéticos).
2. Se construye el modelos con la muestra S.
3. El resto de datos se prueban contra el modelo ajustado. Esos puntos que se ajustan al modelo estimado, de acuerdo con alguna función de pérdida, (conjunto de consenso).
4. El modelo estimado se considera bueno si se han clasificado suficientes puntos como parte del conjunto de consenso. En ese caso, S se toma como formado por *inliers*.

RANSAC

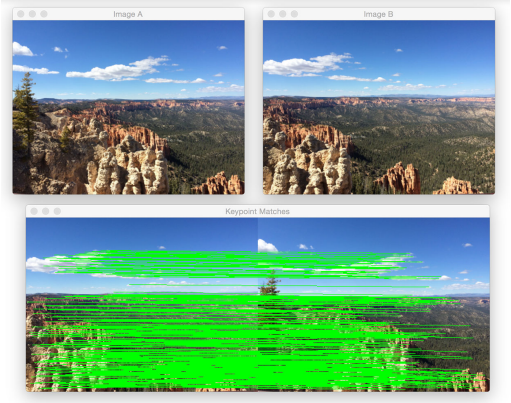


Inliers en RANSAC.

Aplicación: Estimación de homografías en visión computacional.



RANSAC

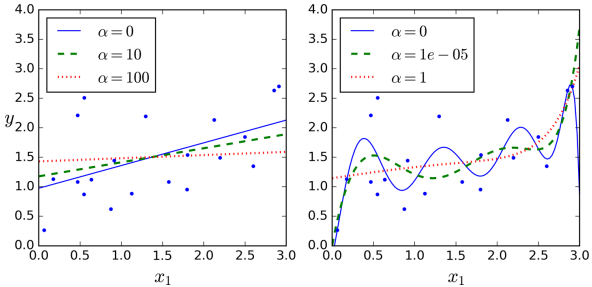


RANSAC: (a) matching de puntos característicos, (b) resultado al hacer *stitching*.

Métodos de Regularización

En ocasiones queremos evitar que nuestro modelo de regresión se ajuste demasiado a los datos (de entrenamiento) y que pierda capacidad para generalizar datos nuevos. En ese caso, es conveniente utilizar **técnicas de regularización**.

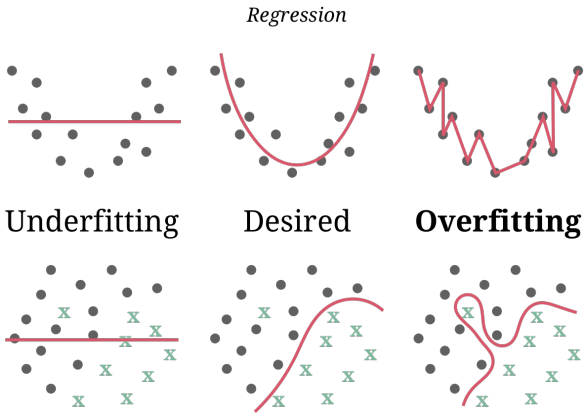
Estos método proporcionan una eficiencia mejorada en problemas de estimación de parámetros, a cambio de una cantidad tolerable de sesgo (*bias-variance tradeoff*).



Regresión polinomial: (a) sin regularización, (b) con regularización.

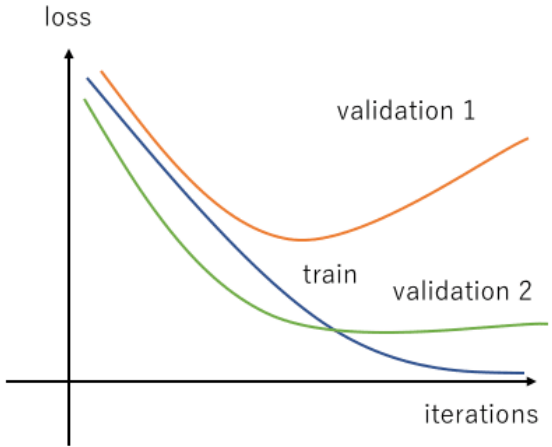
Métodos de Regularización

Sirve para reducir el sobreajuste del modelo a los datos.



Métodos de Regularización

Sirve para reducir el sobreajuste del modelo a los datos.



Regresion Ridge

La **regresión Ridge** o **regularización de Tikhonov**, es un método de regularización para problemas mal planteados. Se utiliza por ejemplo para mitigar el problema de multicolinealidad en la regresión lineal, que ocurre comúnmente en modelos con un gran número de parámetros.

En el caso de regresión lineal ordinaria, se obtiene al modificar el problema de mínimos cuadrados incorporando un término de penalización sobre los coeficientes del vector de parámetros β:

$$J = \sum_{i=1}^{n} (y_i - \mathbf{x}_i^T \beta)^2 + \lambda ||\beta||_2^2,$$

con $\lambda >$ o el parámetro de regularización. En forma vectorial, obtenemos

$$\boxed{(\mathbb{X}^T\mathbb{X} + \lambda I)\beta = \mathbb{X}^T\mathbf{y}} \quad \Rightarrow \quad \widehat{\beta}_R = (\mathbb{X}^T\mathbb{X} + \lambda I)^{-1}\mathbb{X}^T\mathbf{y}.$$

Regresión LASSO

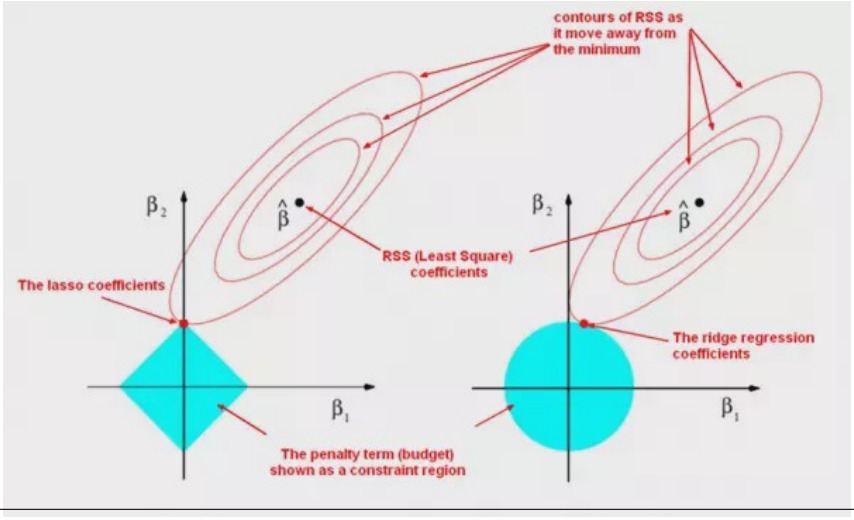
La **regresión LASSO** (*least absolute shrinkage and selection operator*), es un método de regularización y selección de variables. Fue propuesto por Tibshirani (1996).

LASSO se formuló originalmente para modelos de regresión lineal. Por ejemplo, se usa para estimar diferentes alternativas de coeficientes, ya que éstas no necesitan ser únicas si las covariables son colineales.
Se extiende a otros modelos estadísticos, incluidos modelos lineales generalizados, ecuaciones de estimación generalizadas, estimadores M.

En el caso de regresión lineal ordinaria, se obtiene al incorpoar un término de penalización

$$J = \sum_{i=1}^{n} (y_i - \mathbf{x}_i^T \beta)^2 + \lambda ||\beta||_1.$$

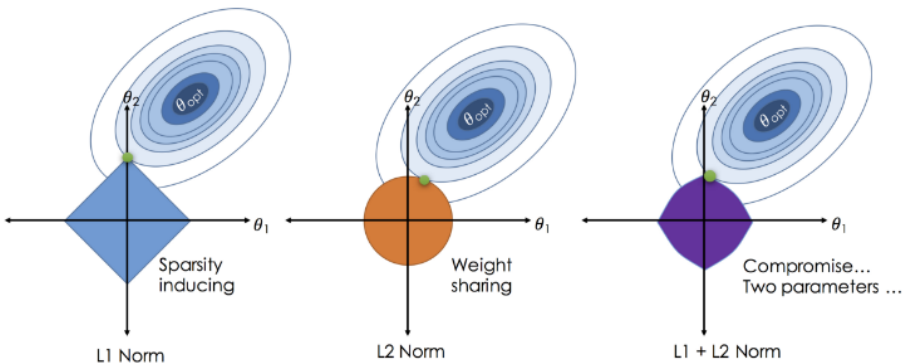
Métodos de Regularización



Métodos de Regularización

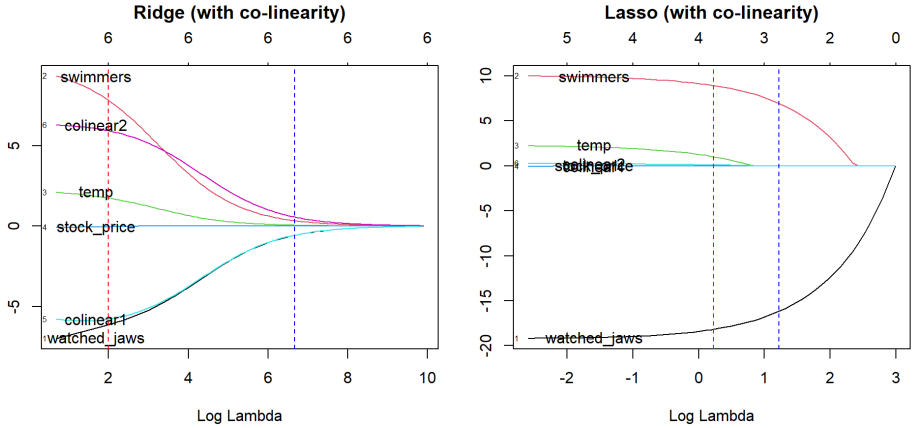
Otro método popular de regularización es el *Elastic Net*:

$$J = \sum_{i=1}^{n} (y_i - \mathbf{x}_i^T \beta)^2 + \lambda_1 ||\beta||_1 + \lambda_2 ||\beta||_2^2.$$



Métodos de Regularización

Los modelos regularizados LASSO inducen raleza (sparcity) sobre el conjunto de coeficientes:



Selección de Variables

Estadísticas:

- coeficientes de correlación (Pearson, Spearman)
- coeficiente τ de Kendall
- ANOVA
- test χ^2
- Información mutua
- score de Fisher
- cotas de varianza

Embedded Methods:

- árboles de decisión (importancias)
- LASSO

Wrapper Methods:

- FFS (*forward feature selection*)
- EFS (*exhaustive feature selection*)
- BFE (*backward feature elimination*)

Selección de Modelos

Es la tarea de seleccionar un modelo estadístico de un conjunto de modelos candidatos. En los casos más simples, se considera un conjunto de datos preexistente, pero también puede implicar el diseño de experimentos de modo que los datos recopilados se adapten bien al problema.

Partimos del principio de Occam (*Occam's razor*): dado un conjunto de datos o fenómeno, entre dos modelos de poder predictivo o explicativo similar, el más probable es el modelo más simple.

Para el caso de modelos de regresión (paramétricos), existen indicadores estadísticos que permiten comparar entre diferentes modelos. Típicamente, éstos miden la eficiencia del modelo en términos de la

$$\text{bondad del ajuste} + \text{complejidad del modelo}.$$

Criterios de Información

- **Akaike information criterion**,
- **Bayesian information criterion**,
- Bridge criterion,
- **Cross-validation**,
- Deviance information criterion,
- False discovery rate,
- Focused information criterion,
- Hannan–Quinn criterion,
- Kashyap information criterion,
- **Likelihood-ratio test**,
- Mallows's *Cp*
- Minimum description length,
- Minimum message length,
- PRESS statistic / criterion,
- Structural risk minimization,
- Stepwise regression,
- Watanabe–Akaike criterion,
- Extended Bayesian Criterion,
- Extended Fisher Criterion.

Criterios de Información

Consideremos un modelo de regresión, con k parámetros, n datos, para el cual $\widehat{L}$ es su máximo valor de verosimilitud.

Criterio de información de Akaike (AIC):

$$\mathrm{AIC} = 2k - 2\log\widehat{L}.$$

Criterio de información bayesiano (BIC):

$$\mathrm{BIC} = k\log n - 2\log\widehat{L}.$$

Criterio de información de Akaike corregido (AICc):

$$\mathrm{AICc} = \mathrm{AIC} + \frac{2k^2 + 2k}{n - k - 1}.$$