Regresión y optimización

Reconocimiento de patrones

Gamaliel Moreno

Enero-julio 2021

UAZ-MCPI

Maestría en Ciencias del Procesamiento de la Información

Table of contents

Regresión lineal

Recordando

- $(\boldsymbol{x}^{(i)}, y^{(i)})$: i-ésimo dato de entrenamiento
- $h_\theta(\boldsymbol{x}^{(i)}))$: predicción de hipótesis h_θ para dato de entrada $\boldsymbol{x}^{(i)})$

$$h_\theta(\boldsymbol{x}) = \sum_{j=0}^{n} \theta_j x_j = \boldsymbol{\theta}^T \boldsymbol{x}$$

  donde asumimos $x_0 = 1$. El número de características es n

- Función cuadrática de costo:

$$J(\boldsymbol{\theta}) = \frac{1}{2} \sum_{i=1}^{m} h_\theta(\boldsymbol{x}^{(i)} - y^{(i)})^2$$

  con m el número de datos en el conjunto de entrenamiento

- Analíticamente: $\boldsymbol{\theta}^* = \underset{\boldsymbol{\theta}}{\operatorname{argmin}}\, J(\boldsymbol{\theta}) = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{y}$

Función cuadrática de costo

- ¿Por qué usamos una función cuadrática de costo?

$$J(\boldsymbol{\theta}) = \frac{1}{2}\sum_{i=1}^{m} h_{\boldsymbol{\theta}}(\boldsymbol{x}^{(i)}) - y^{(i)})^2$$

  como m el número de datos en el conjunto de entrenamiento.
- Pudimos usar otras cosas: valor absoluto , o potencia de 4...
- Vamos a analizar el problema de regresión desde una perspectiva probabilística, como posible argumentación para esto

Regresión lineal

- Supongamos que la salida y las entradas siguen el modelo de regresión

$$y^{(i)} = \boldsymbol{\theta}^T \boldsymbol{x}^{(i)} \epsilon^{(i)} \Rightarrow \epsilon^{(i)} = y^{(i)} - \boldsymbol{\theta}^T \boldsymbol{x}^{(i)}$$

- El término de error $\epsilon^{(i)}$ captura
  - Aspectos no modelos
  - Ruido aleatorio
- Supondremos que $\epsilon^{(i)}$ son independientes e idénticamente distribuidos (i.i.d.)
- Como proceso es complejo, supondremos que $\epsilon^{(i)} \sim \mathcal{N}(0, \sigma^2)$ y por tanto la PDF de $\epsilon^{(i)}$ es

$$p(\epsilon^{(i)}) = \frac{1}{\sqrt{2\pi\sigma}} exp\left(-\frac{(\epsilon^{(i)})^2}{2\sigma^2}\right)$$

- Introduciendo el modelo $\epsilon^{(i)} = y^{(i)} - \boldsymbol{\theta}^T \boldsymbol{x}^{(i)}$ derivamos

$$p(y^{(i)}|\boldsymbol{x}^{(i)}; \boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma}} exp\left(-\frac{(y^{(i)} - \boldsymbol{\theta}^T \boldsymbol{x}^{(i)})^2}{2\sigma^2}\right)$$

Regresión lineal

- Nótese que $\boldsymbol{\theta}$ está fuera de la condición, pues no es una variable aleatoria. La asumimos como dada.
- La distribución $y^{(i)}|\boldsymbol{x}^{(i)};\boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\theta}^T\boldsymbol{x}^{(i)}, \sigma^2)$

Verosimilitud

- Con la matriz de diseño $\boldsymbol{X}$ y los parámetros de $\boldsymbol{\theta}$, la probabilidad conjunta de los datos es $p(\boldsymbol{y}|\boldsymbol{X};\boldsymbol{\theta})$
- $p(\boldsymbol{y}|\boldsymbol{X};\boldsymbol{\theta})$ se interpreta como función de los datos para $\boldsymbol{\theta}$ constante
- Cuando queremos interpretar a $p(\boldsymbol{y}|\boldsymbol{X};\boldsymbol{\theta})$ como función de $\boldsymbol{\theta}$ la llamamos verosimilitud (likelihood)

$$L(\boldsymbol{\theta}) = L(\boldsymbol{\theta}; \boldsymbol{X}, \boldsymbol{y}) = p(\boldsymbol{y}|\boldsymbol{X};\boldsymbol{\theta})$$

- Si suponemos que $\epsilon^{(i)}$ son i.i.d, entonces

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{m} p(\boldsymbol{y}^{(i)}|\boldsymbol{x}^{(i)};\boldsymbol{\theta}) = \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi\sigma}} exp\left(-\frac{(y^{(i)} - \boldsymbol{\theta}^T\boldsymbol{x}^{(i)})^2}{2\sigma^2}\right)$$

Verosimilitud logarítmica

- La verosimilitud logarítmica (log likelihood) es

$$
\begin{aligned}
l(\boldsymbol{\theta}) &= ln(L(\boldsymbol{\theta})) \\
&= ln \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi\sigma}} exp\left(-\frac{(y^{(i)} - \boldsymbol{\theta}^T \boldsymbol{x}^{(i)})^2}{2\sigma^2}\right) \\
&= \sum i = 1^m ln \frac{1}{\sqrt{2\pi\sigma}} exp\left(-\frac{(y^{(i)} - \boldsymbol{\theta}^T \boldsymbol{x}^{(i)})^2}{2\sigma^2}\right) \\
&= mln \frac{1}{2\pi\sigma} - \frac{1}{\sigma^2}\frac{1}{2}\sum_{i}^{m}(y^{(i)} - \boldsymbol{\theta}^T \boldsymbol{x}^{(i)})^2
\end{aligned}
$$

- Observe que maximizar $l(\boldsymbol{\theta})$ eslo mismo que minimizar $J(\boldsymbol{\theta})$
- Con suposiciones probabilíticas: regresión de mínimos cuadrados es equivalente a estimación de máxima verosimilitud de $\boldsymbol{\theta}$ (Note irrelevancia σ)

Selección de características

- En ejemplo de precios de casas, tenemos varias características (features) a disposición:
  - Área habitable
  - Número de pisos
  - Número de habitaciones
- Selección de las cuáles características usar es un criterio de diseños
- Es posible introducir características artificiales para introducir no linealidad en un proceso en principio lineal:

$$\boldsymbol{x} = [1 \quad x_1 \quad x_1^2 \cdots x_1^n]$$

  lo que permite modelar aproximaciones de Taylor de n-ésimo orden, de cualquier función no lineal.

Selección de características

Figures

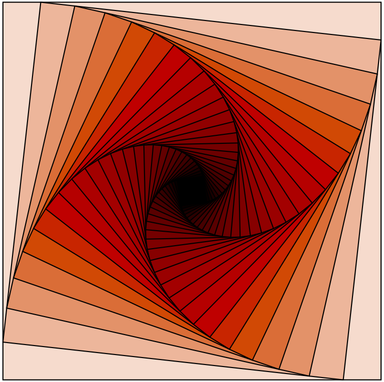


Figure 1: Rotated square from texample.net.

Tables

Table 1: Largest cities in the world (source: Wikipedia)

City	Population
Mexico City	20,116,842
Shanghai	19,210,000
Peking	15,796,450
Istanbul	14,160,467

Blocks

Three different block environments are pre-defined and may be styled with an optional background color.

Default
Block content.

Alert
Block content.

Example
Block content.

Default
Block content.

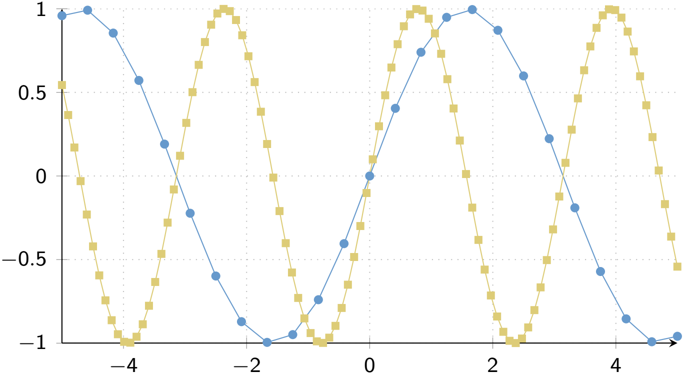
Alert
Block content.

Example
Block content.

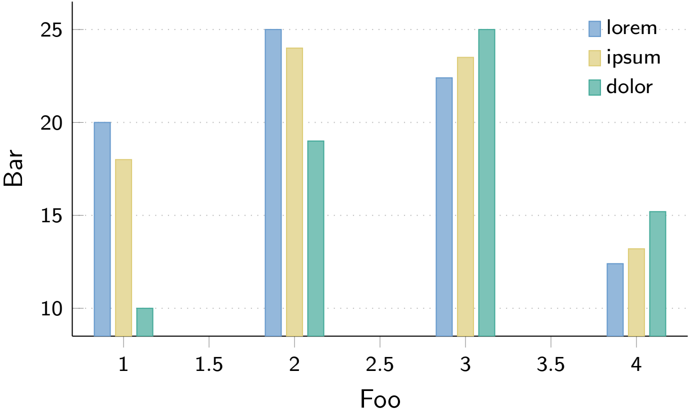
Math

$$e = \lim_{n \to \infty} \left(1 + \frac{1}{n}\right)^n$$

Line plots



Bar charts



Quotes

Veni, Vidi, Vici

Frame footer

metropolis defines a custom beamer template to add a text to the footer. It can be set via

```
\setbeamertemplate{frame footer}{My custom footer}
```

References

Some references to showcase [allowframebreaks] [**?**, **?**, **?**, **?**, **?**]

Conclusion

Summary

Get the source of this theme and the demo presentation from

`github.com/matze/mtheme`

Questions?

Backup slides

Sometimes, it is useful to add slides at the end of your presentation to refer to during audience questions.

The best way to do this is to include the `appendixnumberbeamer` package in your preamble and call `\appendix` before your backup slides.

metropolis will automatically turn off slide numbering and progress bars for slides in the appendix.

References i