Teoría de Probabilidad - Aplicaciones

Cristian Guarnizo-Lemus
cristianguarnizo@itm.edu.co

Maestria en Automatización y Control Industrial

Contenido

Análisis exploratorio de datos (EDA)

En la actualidad, corresponde a un conjunto de técnicas empleadas realizar un análisis inicial de un conjunto de datos. Se hace uso de momentos estadísticos, como la media, varianza, curtosis. También de elementos gráficos de estadística descriptiva, como estimación de densidades (histogramas o KDE), BoxPlots, scatter plots entre otros.

Contenido

Estimación o Inferencia

Proceso por el cual se estima o calcular el valor de variables o parámetros desconocidas. Dentro de los modelos probabilisticos existen

- Estimación puntual: calculo puntual de los valores de los parámetros. Por ejemplo los valores de los parámetros que maximicen la verosimilitud ($p(\mathcal{D}|\theta)$), o el máximo a posteriori ($p(\mathcal{D}|\theta)p(\theta)$).
- Inferencia Bayesiana: calculo del posterior de los parámetros dados las observaciones, $p(\theta|\mathcal{D})$. Si la verosimilitud y el prior son conjugados, entonces el posterior esta distribuido por el mismo tipo de distribución del prior. Si el posterior no es tratable, se recurre a inferencia aproximada usando técnicas de muestreo o variacionales.

Contenido

Estimación Puntual

Consiste en estimar los valores de los parámetros que maximicen una función de probabilidad o un criterio de error.

1. Función de costo: Error medio cuadrático, Verosimilitud, Posterior.
2. Técnicas de optimización: Derivada (pendientes iguales a cero), Gradiente, búsquedas aleatorias.

Contenido

Inferencia Bayesiana

Consiste en determinar el posterior de los parámetros, la cual es la distribucion de los parametros dados los datos

$$p(\theta|\mathcal{D}) \propto p(\mathcal{D}|\theta)p(\theta)$$

1. Prior Conjugado: para ciertas distribuciones de verosimilitud y prior, se puede encontrar que el posterior tiene las misma forma que el prior.
2. Posterior no tratable: Se puede aproxima por medio de técnicas de muestreo: MCMC, MH, HMC, Gibbs sampling. O por métodos variacionales. Para modelos de variable latente se puede emplear el algoritmo EM.

Inferencia Bayesiana - Software

Actualmente existen paquetes de software orientados a estimar distribuciones de variables empleando muestreo, por ejemplo:

- Edward - TensorFlow.
- Pyro - PyTorch.
- MXFusion - MXNet.
- Stan - C++ (R, Python).

Contenido

Contenido

Diagnóstico Médico

Suponga los siguientes eventos $X = 1$ si el resultado del mamograma es positivo, $Y = 1$ si la mujer tiene cancer de mama. Ahora asuma que el test tiene una sensibilidad del 80%, lo que significa, en otras palabras

$$p(X = 1|Y = 1) = 0.8$$

La probabilidad de tener cancer es $p(Y = 1) = 0.004$ y la probabilidad que el test sea positivo dado que la persona no tiene cancer es $p(X = 0|Y = 1) = 0.1$. Calcular la probabilidad que la persona tenga cancer sabiendo que el test salio positivo.

Diagnóstico Médico

Suponga los siguientes eventos $X = 1$ si el resultado del mamograma es positivo, $Y = 1$ si la mujer tiene cancer de mama. Ahora asuma que el test tiene una sensibilidad del 80%, lo que significa, en otras palabras

$$p(X = 1|Y = 1) = 0.8$$

La probabilidad de tener cancer es $p(Y = 1) = 0.004$ y la probabilidad que el test sea positivo dado que la persona no tiene cancer es $p(X = 0|Y = 1) = 0.1$. Calcular la probabilidad que la persona tenga cancer sabiendo que el test salio positivo.

$$p(Y = 1|X = 1) = \frac{p(X = 1|Y = 1)p(Y = 1)}{p(X = 1|Y = 1)p(Y = 1) + p(X = 1|Y = 0)p(Y = 0)}$$

Diagnóstico Médico

Suponga los siguientes eventos $X = 1$ si el resultado del mamograma es positivo, $Y = 1$ si la mujer tiene cancer de mama. Ahora asuma que el test tiene una sensibilidad del 80%, lo que significa, en otras palabras

$$p(X = 1|Y = 1) = 0.8$$

La probabilidad de tener cancer es $p(Y = 1) = 0.004$ y la probabilidad que el test sea positivo dado que la persona no tiene cancer es $p(X = 0|Y = 1) = 0.1$. Calcular la probabilidad que la persona tenga cancer sabiendo que el test salio positivo.

$$p(Y = 1|X = 1) = \frac{p(X = 1|Y = 1)p(Y = 1)}{p(X = 1|Y = 1)p(Y = 1) + p(X = 1|Y = 0)p(Y = 0)}$$

$$p(Y = 1|X = 1) = \frac{0.8 \times 0.004}{0.8 \times 0.004 + 0.1 \times 0.996} = 0.031$$

Contenido

Clasificadores

Podemos generalizar el ejemplo de diagnóstico médico para clasificar vectores de características $\mathbf{x}$

$$p(y = c|\mathbf{x}, \theta) = \frac{p(y = c|\theta)p(\mathbf{x}|y = c, \theta)}{\sum_{c'} p(y = c'|\theta)p(\mathbf{x}|y = c', \theta)}$$

Análisis discriminante Gaussiano

Una aplicación importante de distribuciones normales multivariadas es definir la clase condicional de las densidades en un clasificador, por ejemplo

$$p(\mathbf{x}|y = c, \theta) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c)$$

Si $\boldsymbol{\Sigma}_c$ es diagonal, es equivalente a naive Bayes.
Podemos clasificar un vector de características utilizando la siguiente regla de decisión

$$\hat{y}(\mathbf{x}) = \text{argmax}_c[\log p(y = c|\pi) + \log p(\mathbf{x}|\theta_c)]$$

Contenido

Contenido

Regresión lineal múltiple

Asumiendo que se tienen $\mathbf{y} = [y_1, \ldots, y_N]$ datos que se quieren modelar con la siguiente regresión múltiple

$$y_n = b_0 + \sum_{k=1}^{p} b_k x_{nk} + e_k,$$

donde $e_k \sim \mathcal{N}(0, \sigma^2)$ e iid, representa el error. La expresión anterior se puede escribir de forma matricial

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e},$$

donde $\mathbf{X} \in \mathbb{R}^{N\times(p+1)}$ y $\mathbf{e} \in \mathbb{R}^{N}$.

Regresión lineal múltiple

El modelo queda determinado probabilisticamente así

$$\mathbf{y} \sim \mathcal{N}(\mathbf{Xb}, \Sigma),$$

donde $\Sigma = \sigma^2\mathbf{I}$. A partir de la expresión anterior determinar la solución por máxima verosimilitud de **b**.

Regresión lineal múltiple

La verosimilitud esta dada por

$$p(\mathbf{y}) = \frac{1}{(2\pi)^{k/2}|\mathbf{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{y}-\mathbf{X}\mathbf{b})^\top\mathbf{\Sigma}^{-1}(\mathbf{y}-\mathbf{X}\mathbf{b})\right)$$

El logaritmo de la verosimilitud es

$$\ln p(\mathbf{y}) = -\frac{k}{2}(2\pi) - \frac{1}{2}|\mathbf{\Sigma}| - \frac{1}{2}(\mathbf{y}-\mathbf{X}\mathbf{b})^\top\mathbf{\Sigma}^{-1}(\mathbf{y}-\mathbf{X}\mathbf{b})$$

La derivada con respecto a **b** es

$$\frac{\mathrm{d}\ln p(\mathbf{y})}{\mathrm{d}\,\mathbf{b}} = -2\mathbf{X}^\top\mathbf{\Sigma}^{-1}(\mathbf{y}-\mathbf{X}\mathbf{b})$$

Igualando a cero y despejando **b**

$$\mathbf{b} = (\mathbf{X}^\top\mathbf{\Sigma}^{-1}\mathbf{X})\mathbf{X}^\top\mathbf{\Sigma}^{-1}\mathbf{y}$$

Referencias

- Murphy. "Machine Learning - A Probabilistic Perspective", 2012.
- C. Bishop. "Pattern Recognition and Machine Learning", 2006.