Regression logistique

2 février 2020

Sommaire

Régression

Modèle

- On cherche à prédire une valeur continue $y \in \mathbb{R}$ pour une observation $\mathbf{x} \in \mathbb{R}^d$.
- modèle linéaire

$$y = \mathbf{x}^\top \mathbf{w} + b \quad \text{ou} \quad y = \sum_i w_i h_i(\mathbf{x})$$

y prend des valeurs binaires (en classification)

- ignorer la binarité ?
- considérer le signe de

$$\operatorname{sign}(\mathbf{x}^\top \mathbf{w} + b)$$

Regression pour une probabilité

modèle spécifique

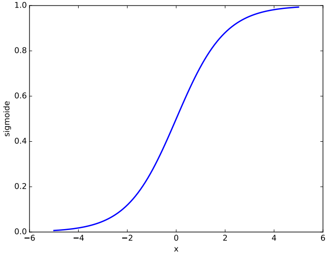
- pour les classes, y $= \{0, 1\}$
- estimer la probabilité de la classe à partir de l'exemple $p(y|\mathbf{x})$, on cherche une sortie $y \in [0, 1]$.

Approximation de la fonction signe

- on utilise un modèle de type **logistique** associé à un modèle linéaire

$$f(\mathbf{x}) = \sigma(\mathbf{w}^\top \mathbf{x} + b)$$

  avec la fonction logistique (ou sigmoide) est $\sigma(\mathbf{z}) = \frac{1}{1+\exp(-\mathbf{z})}$
- fonction dérivable



Interprétation probabiliste de $f(\mathbf{x})$

Modélisation d'une probabilité de classes

- Probabilité comprise entre 0 et 1.
- Modélisation par une fonction logistique

$$p(y = 0|\mathbf{x}) = \sigma(\mathbf{w}^\top \mathbf{x} + b) = \frac{1}{1 + \exp(-\mathbf{w}^\top \mathbf{x} - b)}$$

- Cas d'un problème de classification à deux classes $p(y = 1|\mathbf{x})$?
- comme $y = \{0, 1\}$, on a la propriété suivante :

$$p(y = 0|\mathbf{x}) + p(y = 1|\mathbf{x}) = 1$$

- et

$$p(y = 1|\mathbf{x}) = 1 - \frac{1}{1 + \exp(-\mathbf{w}^\top \mathbf{x} - b)} = \frac{\exp(-(\mathbf{w}^\top \mathbf{x} + b))}{1 + \exp(-(\mathbf{w}^\top \mathbf{x} + b))}$$

- voir illustration

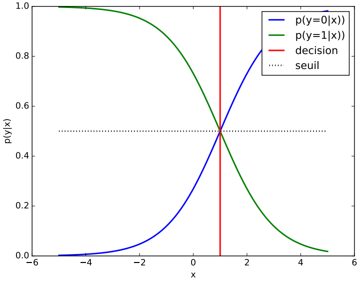
Régression logistique : quel règle de décision ?

Binarisation des probas

- comment décider $y = 0$ ou $y = 1$ à partir des probabilités ?
- on peut décider si $p(y = 1|x) \geq p(y = 0|\mathbf{x})$ ou l'inverse
- comme $p(y = 1|\mathbf{x}) = 1 - p(y = 0|\mathbf{x})$, la frontiere est

$$p(y = 1|\mathbf{x}) = p(y = 0|\mathbf{x}) = 0.5 = \sigma(\mathbf{w}^\top \mathbf{x} + b)$$

  soit $\mathbf{w}^\top \mathbf{x} + b = 0$
- illustration



Apprentissage du modèle de régression logistique ?

Vraisemblance

- Soit des données d'apprentissage $\{\mathbf{x}_i, y_i\}_{i=1}^n$ tirée *i.i.d* d'une loi de proba jointe $P(X, Y)$ avec $\mathbf{x}_i \in \mathbb{R}^d$ $y_i = \{0, 1\}$
- On peut définir la probabilité de l'ensemble de y_i conditionnellement aux $\mathbf{x}_i$ comme

$$p(y_1, y_2, \cdots, y_n | \mathbf{x}_1, \cdots, \mathbf{x}_n)$$

- Si on modélise cette proba par la fonction logistique, la fonction de vraisemblance des paramètres est (où **w** inclut ici **w** et b)

$$L(\mathbf{w}) = p(y_1, y_2, \cdots, y_n | \mathbf{x}_1, \cdots, \mathbf{x}_n; \mathbf{w}) = \prod_{i=1}^{n} p(y_i | \mathbf{x}_i; \mathbf{w})$$

Apprentissage

- Maximisation de la vraisemblance des observations

$$\max_{\mathbf{w}} L(\mathbf{w}) = \max_{\mathbf{w}} \prod_{i=1}^{n} p(y_i | \mathbf{x}_i; \mathbf{w})$$

Optimisation

Cout logistique

- le problème et la fonction objective

$$\max_{\mathbf{w}} L(\mathbf{w}) = \prod_{i=1}^{n} p(y_i|\mathbf{x}_i;\mathbf{w}) = \max_{\mathbf{w}} \prod_{i=1^n} (1 - p(y_i = 0|\mathbf{x}_i))^{y_i} (p(y_i = 0|\mathbf{x}_i)^{1-y_i}$$

car $y_i \in \{0, 1\}$ et $a^0 = 1, \forall\, a$

- minimisation de la log vraisemblance

$$\begin{aligned} L_{log}(\mathbf{w}) &\triangleq -\log L(\mathbf{w}) \qquad (1) \\ &= -\sum_{i=1}^{n} \log(p(y_i|\mathbf{x}_i;\mathbf{w})) \\ &= -\sum_{i=1}^{n} y_i \log(1 - p(y_i = 0|\mathbf{x}_i;\mathbf{w})) - \sum_{i=1}^{n} (1 - y_i) \log(p(y_i = 0|\mathbf{x}_i;\mathbf{w}) \end{aligned}$$

Optimisation par rapport aux paramètres

Simplication du cout

- le cout est

$$L_{log}(\mathbf{w}) = -\sum_{i=1}^{n} y_i \log(\mathbf{p}(y_i = 1|\mathbf{x}_i; \mathbf{w}))) - \sum_{i=1}^{n} (1 - y_i) \log(p(y_i = 0|\mathbf{x}_i; \mathbf{w})$$

avec $p(y = 0|\mathbf{x}) = \frac{1}{1+\exp(-\mathbf{z})}$ et $p(y = 1|\mathbf{x}) = \frac{\exp(-z)}{1+\exp(-\mathbf{z})}$ et $\mathbf{z} = \mathbf{w}^\top \mathbf{x} + b$

- apres simplication

$$L_{log}(\mathbf{w}) = \sum_{i=1}^{n} \log(1 + \exp(-\mathbf{z}_i)) + \sum_{i} y_i \mathbf{z}_i$$

Propriétés

- convexité ?
- condition d'optimalité ?

Algorithme d'optimisation

Solution

- pas de solution analytique
- descente de gradient

Calcul du gradient

$$\frac{\partial L}{\partial w_k} = \sum_i y_i(\mathbf{x}_{i,k}) - \mathbf{x}_{i,k} \frac{\exp(-\mathbf{z}_i)}{1 + \exp(-\mathbf{z}_i)}$$

Descente de gradient

$$\mathbf{w}_k \leftarrow \mathbf{w}_k - \gamma \sum_i \mathbf{x}_{i,k}(y_i - \frac{\exp(-\mathbf{z}_i)}{1 + \exp(-\mathbf{z}_i)})$$

De la régression à la classification

Régression logistique

- modèle linéaire simple
- modéliser les probas des classes
- extension multi-classe naturelle
- algorithme de descente de gradient