Recitation n°2 : MNL

02/10/19

- n products with features $\vec{x}_1, \dots, \vec{x}_n \in \mathbb{R}^d$

<u>Utility</u> : $u_i = \beta^T \vec{x}_i \quad , \; i \in \{1 \dots n\}$

u_i: utility of product i

β: weight of each feature

<u>Randomness</u> : $\tilde{u}_i = \beta^T \vec{x}_i + \tilde{\varepsilon}_i,$

$\tilde{u}_0 = \tilde{\varepsilon}_0$ (outside option)

$\tilde{\varepsilon}_i$: Gumbel $(0, 1)$.

<u>Why Gumbel?</u> a) tractability

b) related to the <u>maximum</u> of samples from a common distribution.

c) $X \sim Exp(1) \Rightarrow -\log(X)$ has distribution $\sim$ Gumb$(0,1)$.

- Which product is chosen?

$$\mathbb{P}_{\tilde{\varepsilon}_1, \dots, \tilde{\varepsilon}_n} (\text{select } j \text{ from } \{1, \dots, n\} \cup \{0\}) = \mathbb{P}_{\tilde{\varepsilon}_1, \dots, \tilde{\varepsilon}_n} (\tilde{u}_j \geq \tilde{u}_i \quad \forall\, i \neq j, i = 0)$$

$$= \pi(j, \{1, \dots, n\} \cup \{0\})$$

Using $\tilde{\varepsilon}_i \sim$ Gumbel $(0, 1)$, and (b),

$$\pi(j, \{1, \dots, n\} \cup \{0\}) = \frac{e^{\beta^T x_j}}{1 + \sum_{i=1}^{n} e^{\beta^T x_i}}$$

Think about ...

→ Category feature: how to model "color"?

color $\in$ {red, blue, yellow} → introduce more dimensions for features and new 0/1 values.

→ Time-Dependency: if each product i has features x_i^t, $t = 1, \dots, T$, we can reduce the new problem to the stationary case by extending the dimensions to $d \times T$:

$$u_i = (\beta, \dots, \beta)^T (x_i^1, x_i^2, \dots, x_i^T) + \tilde{\varepsilon}_i, \; i = 1 \dots n.$$

How to fit the Data:

We have observed $(S_1, j_1), \dots, (S_k, j_k)$.

subset at time 1

product selected at time 1

MLE: Maximum Likelihood Estimation

$$\mathbb{P}(\text{observing } (S_1, j_1), \dots, (S_k, j_k)) = \prod_{i=1}^{k} \mathbb{P}(\text{observing } (S_i, j_i)) = \prod_{i=1}^{k} \frac{e^{\beta^T x_{j_k}}}{1 + \sum_{i \in S_k} e^{\beta^T x_i}}$$

independence of $\tilde{\varepsilon}_i$

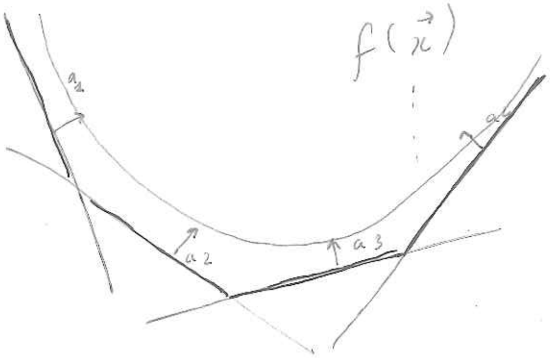
$\ulcorner$B

Using the log-likelihood, we want

$$\max_{\vec{\beta}} \sum_{\ell=1}^{k} \log\left(\frac{e^{\beta^T x_{j_\ell}}}{1+\sum_{i\in S_\ell} e^{\beta^T x_i}}\right) = \max_{\vec{\beta}} \sum_{\ell=1}^{k} \beta^T x_{j_\ell} - \log\left(1+\sum_{i\in S_\ell} e^{\beta^T x_i}\right)$$

$\boxed{\text{Note}}$ $f: x \mapsto \log\left(\sum_{j=1}^{k} e^{a_j^T x}\right)$ is "soft-max"

it approximates (in a smooth way) $x \mapsto \max\{a_1^T x, \dots, a_k^T x\}$



Question: Choice of $S = \{1 \dots n\}$ among millions of products?

→ «Word-to-vec»: create a vector that counts for the occurences of the word of the text; create clusters / Nearest-Neighbors.

Given S, Find prices $p_1, \dots, p_n$ to maximize revenue, where $\vec{x}_j = (p_j, x_{j2}, \dots, x_{jd})$

$$u_j = -\beta_1 \cdot p_j + \underbrace{\sum_{\ell=2}^{d} \beta_\ell x_{j\ell}}_{\alpha_j}.$$

Γ_c

One can prove:

$$\left| \begin{array}{l} p_j^* = \dfrac{1}{\beta_1} \\ j \in \{1 \dots n\} \end{array} \right. \quad \text{is optimal for} \quad \max \sum_{j=1}^{n} p_j \frac{e^{\alpha_j - \beta_1 p_j}}{1 + \sum_{i \in S} e^{\alpha_i - \beta_1 p_i}}$$

(note uniform price)

$$\left| \begin{array}{l} p_j^* = \dfrac{1}{\beta_1} + c_j \\ j \in \{1 \dots n\} \end{array} \right. \quad \text{is optimal for} \quad \max \sum_{j=1}^{n} (p_j - c_j) \cdot \frac{e^{\alpha_j - \beta_1 \cdot p_j}}{1 + \sum_{i=1}^{n} e^{\alpha_i - \beta_1 p_i}}$$

c_j: price of product j.