Generativní modely dat popsané stromovou strukturou

Jakub Bureš

Katedra matematiky

Vedoucí práce: doc. Ing Václav Šmídl, Ph.D.

25. srpna 2020

Přehled

Generativní modely

Generativní model

Mějme nějakou množinu datových záznamů $\mathbf{x} = \{x_1, \ldots, x_n\}$, představující nezávislé proměnné a nějakou množinu $\mathbf{y} = \{y_1, \ldots, y_n\}$, jakožto závislé proměnné. Generativní model je potom takový model, který se učí sdruženou hustotu pravděpodobnosti $p(x, y)$.

- Odhad hustoty pravděpodobnosti $p(x, y)$.
- Dva přístupy:
  1. Rozklad sdružené hustoty pomocí součinového pravidla
  2. Variační autoencoder $\Rightarrow$ ELBO
- Odhad sdružené hustoty pravděpodobnosti stromových struktur.

Odhad hustoty pravděpodobnosti

1. Součinové pravidlo $p(y, x) = p(y|x) \cdot p(x)$
   - Problém převeden na hledání dvou hustot
   - $p(x)$ určíme pomocí maximálně věrohodného odhadu nebo histogramu.
   - $p(y|x)$ určíme pomocí metody nejmenších čtverců pro model

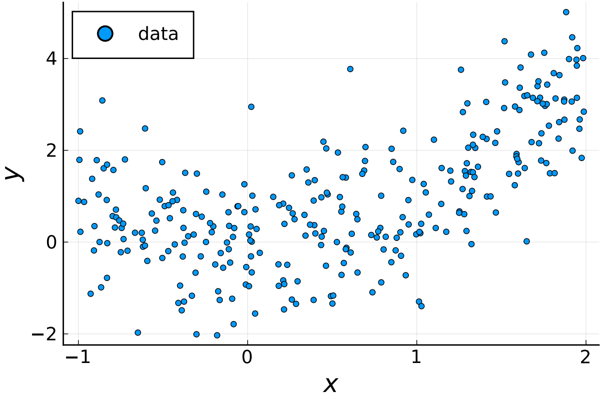
$$\mathbf{y} = \mathbb{X} \cdot \theta + \epsilon. \tag{1}$$

Předpokládáme $\varepsilon_i \sim \mathcal{N}\left(0, \sigma^2\right)$ a položme $X^{\mathsf{T}} = \left(1\ x\ x^2 \ldots x^s\right)$, kde s je stupeň polynomu, jakým data prokládáme. Obdržíme tvar hustoty

$$p(y|x) = \mathcal{N}\left(X^{\mathsf{T}} \cdot \theta, \sigma^2\right) \tag{2}$$

Odhad hustoty pravděpodobnosti

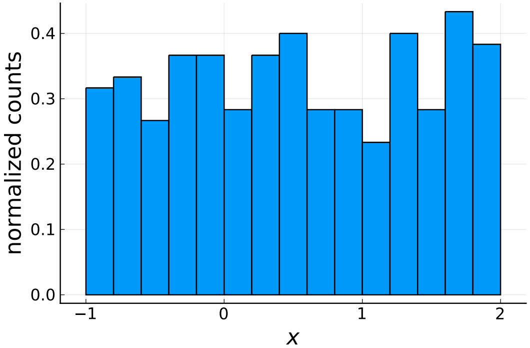
Pro demonstraci problému využijeme následující data



Obrázek: Zadaná data, ze kterých hledáme sdruženou hustotu $p(x, y)$.

Odhad hustoty pravděpodobnosti

Nejprve zobrazíme histogram x–ových složek



Obrázek: Histogram x–ových složek zadaných dat.

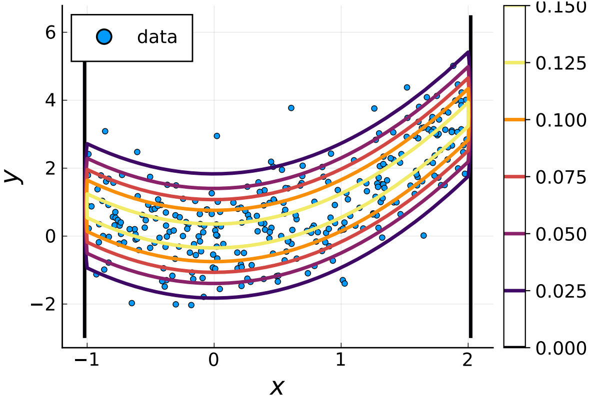
Histogram odpovídá normálnímu rozdělení

$$p(x) = \mathrm{U}\left(-1, 2\right) \tag{3}$$

Odhad hustoty pravděpodobnosti

Podmíněnou hustotu pravděpodobnosti určíme podle vztahu (2). Pro sdruženou hustotu $p(y, x)$ pak dostaneme vztah

$$p(y, x) = \mathrm{U}\,(-1, 2) \cdot \mathcal{N}\left(X^{\mathsf{T}} \cdot \theta, \sigma^2\right) \tag{4}$$



Variační autoencoder

Cílem je najít distribuce $p(\mathbf{x})$ vzorků $\{x_i\}_{i=1}^{n}$. Předpokládáme následující vztahy

$$\mathbf{x} = f_\theta(\mathbf{z}) + \epsilon, \tag{5}$$

kde $\epsilon \sim \mathcal{N}\left(0, \sigma^2 \cdot \mathbb{I}\right)$ a $f_\theta(\mathbf{z})$ je neznámá transformace s parametry θ, kterou se chceme naučit. Využijeme následující formu aproximace

$$p(\mathbf{x}) = \int p(\mathbf{x}|\mathbf{z})p(\mathbf{z})\mathrm{d}\mathbf{z}. \tag{6}$$

Podle vztahu pro **x** určíme distribuce $p\left(\mathbf{x}|\mathbf{z}\right)$ a $p\left(\mathbf{z}\right)$ zvolíme jednoduše

$$\begin{aligned} p(\mathbf{x}|\mathbf{z}) &= \mathcal{N}\left(f_\theta(\mathbf{z}), \sigma^2 \cdot \mathbb{I}\right), \\ p(\mathbf{z}) &= \mathcal{N}\left(0, \mathbb{I}\right). \end{aligned} \tag{7}$$

Variační autoencoder

Víme-li, že

$$D_{KL}(q\|p) = \int q(x)\log\frac{q(x)}{p(x)}\mathrm{d}x, \tag{8}$$

můžeme použít ELBO.

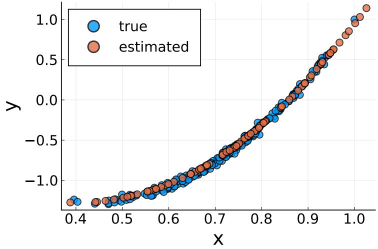
$$\begin{aligned} D_{KL}\left(q\left(\mathbf{z}|\mathbf{x}\right)\|p(\mathbf{z}|\mathbf{x})\right) &= \mathbb{E}_q\left[\log q(\mathbf{z}|\mathbf{x}) - \log p\left(\mathbf{z}|\mathbf{x}\right)\right] \\ &= \mathbb{E}_q\left[\log q(\mathbf{z}|\mathbf{x}) - \log p(\mathbf{x}|\mathbf{z}) - \log p(\mathbf{z}) + \log p(\mathbf{x})\right]. \end{aligned} \tag{9}$$

Tuto rovnici můžeme přepsat pomocí další KL–divergence

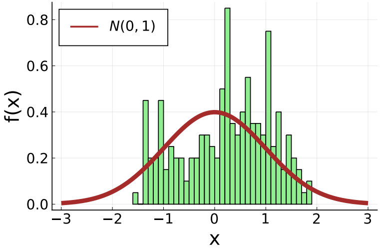
$$\log p(\mathbf{x}) - D_{KL}\left(q\left(\mathbf{z}|\mathbf{x}\right)\|p(\mathbf{z}|\mathbf{x})\right) = \mathbb{E}_q\left[\log p(\mathbf{x}|\mathbf{z})\right] - D_{KL}\left(q\left(\mathbf{z}|\mathbf{x}\right)\|p(\mathbf{z})\right), \tag{10}$$

kde pravá strana této rovnice je lower bound objektu $\log p(\mathbf{x})$, tedy ELBO.

Variační autoencoder



(a) Skutečné vzorky $\{\mathbf{x}, \mathbf{y}\}$ (modře) a jejich odhad pomocí VAE (červeně).



(b) Histogram vzorků $\mathbf{z}$, určených pomocí pseudoinverzní transformační funkce.

- Červené vzorky byly vygenerovány pomocí $f_\theta(z) : \mathbb{R}^1 \to \mathbb{R}^2$
- Histogram byl vygenerován pomocí $g(x, y) : \mathbb{R}^2 \to \mathbb{R}^1$

Stromové struktury

Příklad inspirovaný finanční aplikací.

- Dvě gaussovské směsi (GM)

$$
\begin{aligned}
p(x_i|y=1) &= w_1 \cdot \mathcal{N}\left(\mu_1, \sigma_1^2\right) + (1-w_1) \cdot \mathcal{N}\left(\mu_2, \sigma_2^2\right), \\
p(x_j|y=0) &= w_2 \cdot \mathcal{N}\left(\mu_3, \sigma_3^2\right) + (1-w_2) \cdot \mathcal{N}\left(\mu_4, \sigma_4^2\right),
\end{aligned} \tag{11}
$$

  kde $x_i|y=1$ pro $i \in \hat{a}$, značí hodnotu transakce na bankovních účtech klientů, kteří jsou schopni splácet půjčku a $x_j|y=0$ pro $j \in \hat{b}$, značí totéž, pouze pro ty, kteří nejsou schopni splácet půjčku.

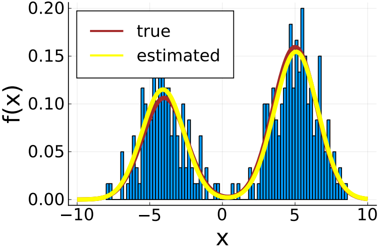
- Dále uvažujeme počet transakcí N_x za sledované období na účtu jednoho klienta pro jednotlivé třídy

$$
\begin{aligned}
p(N_{\mathrm{kx}}^{(1)}|y=1) &= \mathrm{Po}\left(\lambda_1\right) \quad k \in \hat{c}, \\
p(N_{\mathrm{mx}}^{(0)}|y=0) &= \mathrm{Po}\left(\lambda_2\right) \quad m \in \hat{d}.
\end{aligned} \tag{12}
$$

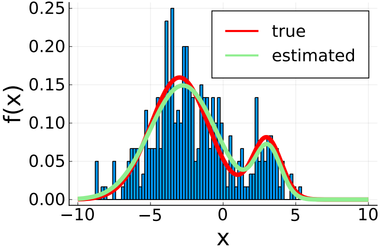
Stromové struktury

- Jak nyní rozhodnout, do které třídy (schopný či neschopný splácet) patří nový klient?
- Nejprve odhadneme všechny parametry všech rozdělení pomocí MLE.

$$
\begin{aligned}
\hat{\mu_1}, \hat{\mu_2}, \hat{\sigma_1}^2, \hat{\sigma_2}^2, \hat{w}_1 &= \underset{\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,w_1}{\arg\max} \log\left(\prod_{i=1}^{a} p(x_i|y=1)\right) \\
\hat{\mu_3}, \hat{\mu_4}, \hat{\sigma_3}^2, \hat{\sigma_4}^2, \hat{w}_2 &= \underset{\mu_3,\mu_4,\sigma_3^2,\sigma_4^2,w_2}{\arg\max} \log\left(\prod_{j=1}^{b} p(x_j|y=0)\right), \\
\hat{\lambda}_1 &= \frac{1}{c}\sum_{k=1}^{c} N_{\mathrm{kx}}^{(1)}, \\
\hat{\lambda}_0 &= \frac{1}{d}\sum_{m=1}^{c} N_{\mathrm{mx}}^{(0)}
\end{aligned}
\tag{13}
$$



(c) $p(x|y=1)$



(d) $p(x|y=0)$

Obrázek: Dvě GM, kde červenou a hnědou barvou jsou nakresleny skutečné distribuce, zeleně a žlutě jsou jejich MLE.

- MLE zde funguje i přesto, že MLE gaussovské směsi nemá analytické řešení.

Stromové struktury

- V dalším kroku sestavíme

$$\begin{aligned} p\left(\mathbf{x}, N_x^{(1)} | y = 1\right) &= \left(\prod_{i=1}^{N_x^{(1)}} p(x_i | y = 1)\right) \cdot p\left(N_x^{(1)} | y = 1\right), \\ p\left(\mathbf{x}, N_x^{(0)} | y = 0\right) &= \left(\prod_{j=1}^{N_x^{(0)}} p(x_j | y = 0)\right) \cdot p\left(N_x^{(0)} | y = 0\right) \end{aligned} \tag{14}$$

- S jejich pomocí provedeme testy poměrem věrohodností

$$\begin{aligned} \Lambda_0(\mathbf{x}) &= \frac{p\left(\mathbf{x}, N_x^{(0)} | y = 0\right)}{p\left(\mathbf{x}, N_x^{(1)} | y = 1\right) + p\left(\mathbf{x}, N_x^{(0)} | y = 0\right)} \in \langle 0, 1 \rangle \\ \Lambda_1(\mathbf{x}) &= \frac{p\left(\mathbf{x}, N_x^{(1)} | y = 1\right)}{p\left(\mathbf{x}, N_x^{(1)} | y = 1\right) + p\left(\mathbf{x}, N_x^{(0)} | y = 0\right)} \in \langle 0, 1 \rangle. \end{aligned} \tag{15}$$

Stromové struktury

- Může se stát, že klient dobře nezapadne do žádné ze skupin
- Test můžeme vylepšit a přidat třetí třídu $y = 2$, v takovém případě dostaneme test

$$\Lambda_2(\mathbf{x}) = \frac{p(\mathbf{x}, N_x^{(2)}|y = 2)}{p(\mathbf{x}, N_x^{(2)}|y = 2) + p(\mathbf{x}, N_x^{(1)}|y = 1) + p(\mathbf{x}, N_x^{(0)}|y = 0)} \tag{16}$$

Děkuji za pozornost

BISHOP, Christopher M. : *Pattern recognition and machine learning*. [New York]: Springer, 2006. Information science and statistics. ISBN 0-387-31073-8.

MANDLÍK, Šimon. *Mapování internetu - modelování interakcí entit v komplexních heterogenních sítích*. Praha, 2020. Diplomová práce. České vysoké učení technické v Praze. Výpočetní a informační centrum.

JIROVSKÝ, Lukáš. *Teorie grafů ve výuce na střední škole*. Praha, 2008. Diplomová práce. Univerzita Karlova, Matematicko-fyzikální fakulta.

PEVNÝ, Tomáš a Petr SOMOL. Discriminative models for multi-instance problems with tree structure. *Proceedings of the 2016 ACM Workshop on Artificial Intelligence and Security*. 2016, 83–91.

SHAFKAT, Irhum. *Intuitively understanding variational autoencoders* [online]. [cit. 2020-07-22]. Dostupné z:

https://towardsdatascience.com/intuitivelyunderstanding-variational-autoencoders-1

RUDER, Sebastian. An overview of gradient descent optimization algorithms. *ArXiv preprint arXiv:1609.04747.* 2016, 1–14.

KINGMA, Diederik P. a Max WELLING. Auto-encoding variational bayes. ArXiv preprint arXiv:1312.6114. 2013, 1–14.

KOVÁŘ, Jan a Niels VAN DE MEER. *Zápisky z míry a pravděpodobnosti.* Praha, 2020. Vysokoškolská skripta. Fakulta jaderná a fyzikálně inženýrská ČVUT v Praze.

KŮS, Václav a Martin KOVANDA.*Matematická statistika.* Praha, 2020. Vysokoškolská skripta. Fakulta jaderná a fyzikálně inženýrská ČVUT v Praze.

LEARNED-MILLER, Eric. *Vector, Matrix, and Tensor Derivatives* [online]. [cit. 2020-07-22]. Dostupné z: http://cs231n.stanford.edu/vecDerivs.pdf

- JEFFREYS, Harold. An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences.* 1946, 1–9.
- JORDAN, Michael Irwin a Andrew Y NG. Advances in neural information processing systems: On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes. 2002.
- ŠMÍDL, Václav. *Linear Regression, Automatic Relevance Determination* [online]. Czech Academy of Sciences [cit. 2020-07-22]. Dostupné z: `http://staff.utia.cas.cz/smidl/files/hbm2020/prezentace03_20.pdf`
- COMMENGES, Daniel. Information Theory and Statistics: an overview. *ArXiv preprint arXiv:1511.00860.* 2015, 1–22.
- PEVNÝ, Tomáš a Marek DĚDIČ. Nested Multiple Instance Learning in Modelling of HTTP network traffic.*ArXiv preprint arXiv:2002.04059.* 2020, 1–13.

BEZANSON, Jeff, Stefan KARPINSKI, Viral B. SHAH a Alan EDELMAN. Julia: A fast dynamic language for technical computing. *ArXiv preprint arXiv:1209.5145.* 2012, 1–27.

YANG, Xitong. *Understanding the Variational Lower Bound* [online]. 2017 [cit. 2020-07-23]. Dostupné z: http://legacydirs.umiacs.umd.edu/~xyang35/files/understanding-variational-lower.pdf