UROB 2023
Midterm test
Varianta: A

Jméno: Yauheni Zviazdau

Body 12,3

1. **MLE regrese**: Máte na střeše senzor, který měří fyzikální veličinu $x \in \mathbb{R}^+$ související s rychlostí větru, jako je např. úhlová rychlost větrného mlýnku. Předpokládáme, že pro dané měření x, má rychlost větru $y \in \mathbb{R}^+$ Rayleighovo rozdělení pravděpodobnosti:

$$p(y|x, w) = (y - wx) \exp\big(-(y - wx)^2\big)$$

kde $w \in \mathbb{R}$ je neznámý parametr.

- 1.1 Dostanete trénovací sadu $\mathcal{D} = \{(x_1, y_1) \dots (x_N, y_N)\}$ měřených veličin x_i a odpovídajících rychlostí větru y_i. Zapište optimalizační problém, který odpovídá maximálně věrohodnému odhadu parametrů modelu w a pokud je to možné, zjednodušte výsledný optimalizační problém na vhodnou **ztrátovou funkci**.

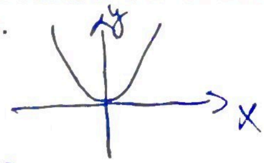
MLE: $\max \prod p(y|x,w)$ ✓ 1

$$w^* = \arg\max_w \prod p(y|x,w) = \arg\max_w \prod (y_i - wx_i) e^{-(y_i - wx_i)^2} = \quad ✗$$

$$= \arg\max_w \sum (y_i - wx_i) - (y_i - wx_i)^2 = \arg\max \sum y_i - wx_i - y_i^2 + 2y_i w x_i - w^2 x_i^2 =$$

$$= \arg\max \sum y_i - y_i^2 + \sum 2 y_i w x_i - wx_i - w^2 x_i^2$$

- 1.2 **Nakreslete výpočetní graf** odvozené ztrátové funkce pro jeden trénovací příklad (x, y) a spočítejte jeho **gradient**.



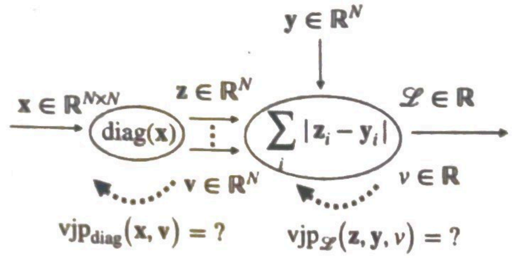
✗ ✗ 0

grad $= 2yx - x - 2wx^2$

- 1.3 Uvažujte nyní, že 1/10 měření nesouvisí se skutečnou rychlostí větru (např. kvůli nějaké vnitřní poruše senzoru). Proto v 1/10 všech tréninkových příkladů pochází rychlost větru y z rovnoměrného rozdělení $y \sim U(0, y_{max})$. **Nakreslete tvar rozdělení pravděpodobnosti**, které modeluje takový případ.

✗ 0

2. **Vector-jacobian-product:** Uvažujte výpočetní graf níže:



Graf se skládá ze dvou funkcí:

a) diag($\mathbf{x}$) funkce, která vrací úhlopříčku vstupní matice $\mathbf{x} \in \mathbb{R}^{N\times N}$ jako sloupcový vektor $\mathbf{y} \in \mathbb{R}^{N}$, který je složený z diagonálních prvků. Například pro $N = 3$, funkce funguje následovně:

$$\mathbf{y} = \mathrm{diag}(\mathbf{x}) = \mathrm{diag}\left(\begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix}\right) = \begin{bmatrix} x_{11} \\ x_{22} \\ x_{33} \end{bmatrix}$$

b) L1-ztrátová funkce, která je definována jako součet absolutních hodnot rozdílů:

$$\mathcal{L} = \sum_i |\mathbf{z}_i - \mathbf{y}_i|$$

- 2.1 Navrhněte efektivní implementaci funkce vracející **vector-jacobian-product** $\mathrm{vjp}_{\mathrm{diag}}(\mathbf{x}, \mathbf{v})$, kde vektor $\mathbf{v} \in \mathbb{R}^N$ je upstream gradient.

Funkce má rozmístit prvky $\vec{v}$ na diagonále nulové matice

✓ 2

- 2.2 Vypočtěte **gradient** ztrátové funkce $\mathcal{L}$ vzhledem k $\mathbf{x}$ (tj. $N \times N$ matice).

$$\frac{\partial L}{\partial x_{11}} = \frac{\partial L}{\partial y_1} \cdot \frac{\partial y_1}{\partial x_{11}} \quad \leftarrow y_1 = x_{11} \Rightarrow \frac{\partial y_1}{\partial x_{11}} = 1$$

$$\text{grad}(|z_i - y_i|) = \begin{cases} -1, & z_i - y_i > \\ 1, & z_i - y_i < \end{cases}$$

$$\frac{\partial L}{\partial x_{N,N}} = \frac{\partial L}{\partial y_N} \cdot \underset{\overset{\shortparallel}{1}}{\frac{\partial y_N}{\partial x_{N,N}}} \quad \leftarrow y_N = x_{N,N} \Rightarrow \frac{\partial y_N}{\partial x_{N,N}} = 1$$

$$\frac{\partial L}{\partial x} = \begin{bmatrix} -\text{sign}(z_1 - y_1) & 0 & \cdots & 0 \\ 0 & -\text{sign}(z_2 - y_2) & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & -\text{sign}(z_N - y_N) \end{bmatrix}$$

✓ 2

- 2.3 Diskutujte rozdíl mezi vaší efektivní implementací $\text{vjp}_{\textbf{diag}}$ a naivním násobením jakobiánů (pouze jedna krátká věta).

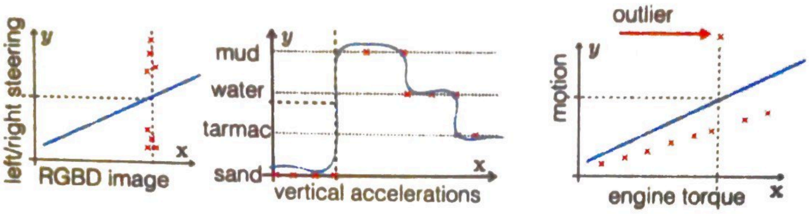
Efektivní implementace bude rychleji a snadněji

✓ 1

3. **Losses and Overfitting:**

- 3.1 Během přednášek jsme diskutovali následující tři problémy, které se objevují při fitování funkce do dat ve smyslu nejmenších čtverců. **Jaký je hlavní zdroj, který tyto problémy způsobuje?** Vyberte právě jednu odpověď.

0



  - overfitting
  - underfitting
  - výsledný problém je nepříznivý pro optimalizaci (velké plochy s nulovým gradientem)
  - odpovídající tvar ztrátové funkce má ostré minimum, které způsobuje špatné zobecnění testovacích dat
  - model předpokládá, že výstupy y mají normální (Gaussovské) rozdělení pro danou hodnotou **x**.
  - KL-divergence není definována
  - je velký rozdíl mezi trénovacím/testovacím pravděpodobnostním rozdělením (training/testing distribution mismatch).

- 3.2 **Jak mohu omezit přefitovaní (overfitting)?** Vyberte žádnou, jednu nebo více odpovědí:

0,3

  - Nikdy nepředpokládat Gaussovský šum.
  - Použití hodně velké (ideálně nekonečné) datové sady
  - Vyhnout se plochému minimu (flat minimum) ve ztrátové funci
  - Musím udržovat váhy v klasifikátoru/regresoru vždy kladné.
  - Musím udržovat váhy v klasifikátoru/regresoru vždy záporné.
  - Musím použít co nejhlubší konvoluční sítě.
  - Musím použít správný model, který zahrnuje co nejvíc apriorních znalostí o řešeném problému.
  - Vyhnout se ostrému minimu (sharp minimum) ve ztrátové funci.
  - Použít plně diferencovatelnou ztrátovou funkci.
  - Implementovat všechny Vector-Jacobian-Product funkce na GPU.
  - Vždy předpokládat Gaussovský šum.

3,5

- 3.3 Označte, které z následujících tvrzení jsou PRAVDA/NEPRAVDA:

PRAVDA ✓ – 2D fitování přímky ve smyslu nejmenších čtverců odpovídá minimalizaci KL-divergence mezi skutečným rozdělením dat p_{data} a normálním (Gaussovským) rozdělením s lineárním průměrem zkonstruovaným následovně

$$p(\mathbf{x}, y|\mathbf{w}) = \mathcal{N}(y;\ w_1 x + w_0,\ \sigma^2)$$

NEPRAVDA ✓ – Hodnota globálního optima následujícího problému je vždy rovna nule

$$\min_{\mathbf{w}} D_{KL}\big(p_{\text{data}}(\mathbf{x}, y) \parallel p(\mathbf{x}, y|\mathbf{w})\big) = 0$$

PRAVDA ✓ – Hodnota globálního optima následujícího problému je vždy nezáporná

$$\min_{\mathbf{w}} D_{KL}\big(p_{\text{data}}(\mathbf{x}, y) \parallel p(\mathbf{x}, y|\mathbf{w})\big) \geq 0$$

NEPRAVDA ✓ – Konvoluční síť se nikdy nepřefitovává na obrazových datech.

NEPRAVDA ✓ – $D_{KL}\big(p_{\text{data}}(\mathbf{x}, y) \parallel p(\mathbf{x}, y|\mathbf{w})\big)$ je monotónně rostoucí funkce vah $\mathbf{w}$.

NEPRAVDA ✓ – $D_{KL}\big(p_{\text{data}}(\mathbf{x}, y) \parallel p(\mathbf{x}, y|\mathbf{w})\big)$ má právě jedno lokální minimum.

PRAVDA ✓ – Gradient konvoluční vrstvy lze spočítat jako konvoluci.

PRAVDA × – Každý vstup do konvoluční vrstvy vždy ovlivňuje všechny pixely ve výstupní příznakové mapě (feature map).

$$M = \left\lfloor \frac{N + 2\,pad - K}{stride} + 1 \right\rfloor$$

4. **Convolution:** Předpokládejme, že vstupní příznaková mapa (feature map) do konvoluční vrstvy je $3 \times 10 \times 10$. in

- 4.1 Jaká je velikost a počet konvolučních jader/filtrů (kernels/filters), pokud by výstup měl být $10 \times 3 \times 3$? (předpokládejme stride=1, padding=0 a dilation_rate=1)

1

$$3 = \left\lfloor \frac{10 + 2 \cdot 0 - K}{1} \right\rfloor + 1 = 10 - K + 1 \; ; \; K = 8.$$

k_size = $3 \times 8 \times 8$ ✓

k_number = 10 ✓

- 4.2 Jaký je padding konvoluční vrstvy, pokud je velikost jádra/filtru $3 \times 3 \times 3$ (kern) a výstup by měl být $7 \times 12 \times 12$ (out) (stride=1 a dilation_rate = 1)?

1

$$12 = \left\lfloor \frac{10 + 2\,pad - 3}{1} \right\rfloor + 1 = 8 + 2\,pad$$

pad = 2 ✓

- 4.3 Uvažujme síť sestávající se ze dvou konvolučních vrstev (bez jakékoli aktivační funkce). Každá vrstva má pouze jedno jádro 3×3 následujícího tvaru:

0,5

$$\mathbf{w} = \begin{bmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \\ w_{31} & w_{32} & w_{33} \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{bmatrix},$$

Můžete tyto dvě vrstvy nahradit plně propojenou vrstvou (fully-connected layer)? Pokud je to možné, jaké budou její váhy? Pokud to není možné, zdůvodněte.

Je to možný ✓

Př img $= 3 \times 3$ $W = 2 \times 2$ $V = 2 \times 2$ $W = \begin{bmatrix} w_{11} & w_{12} \\ w_{2,1} & w_{22} \end{bmatrix}$

$$img = \begin{bmatrix} x_{1,1} & x_{1,2} & x_{1,3} \\ x_{2,1} & x_{2,2} & x_{2,3} \\ x_{3,1} & x_{3,2} & x_{3,3} \end{bmatrix}$$

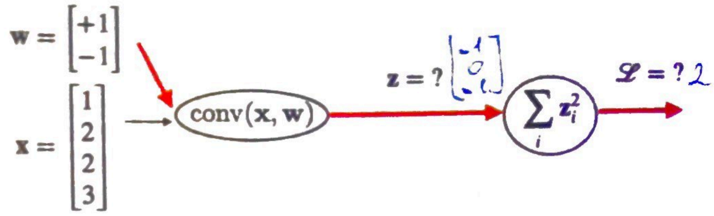
$$conv(img, W) \begin{bmatrix} x w_{11} + x w_{12} + x_{21} w_{21} + w_{22} x_{22} & x_{1,2} w_{11} + x_{13} w_{12} + x_{22} w_{21} + x_{23} w_{22} \\ x_{21} w_{21} + x_{22} w_{12} + x_{3,1} w_{21} + x_{3,2} w_{22} & x_{2,2} w_{11} + x_{2,3} w_{1,2} + x_{3,2} w_{2,1} + x_{33} w_{22} \end{bmatrix}$$

conv(img, V) =

+ + + +

- 4.4 Uvažte následující výpočetní graf s 1D konvolucí. Vypočtěte $\frac{\partial \mathbf{z}}{\partial \mathbf{w}}$ a $\frac{\partial \mathcal{L}}{\partial \mathbf{w}}$:

0



$$z = \begin{bmatrix} -1 \\ 0 \\ -1 \end{bmatrix} \qquad \mathcal{L} = 1+0+1=2$$

$$\frac{\partial \mathcal{L}}{\partial w} = \frac{\partial \mathcal{L}}{\partial z} \cdot \frac{\partial z}{\partial w}$$

$$\frac{\partial \mathcal{L}}{\partial z} = \begin{bmatrix} 2z_1 \\ 2z_2 \\ \vdots \\ 2z_n \end{bmatrix} \quad \times$$

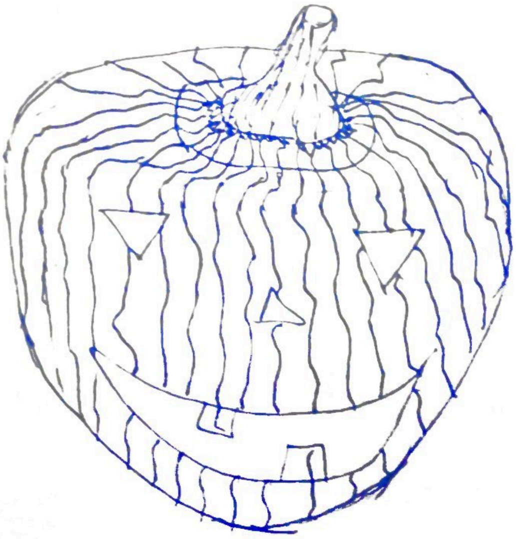
$\frac{\partial \mathbf{z}}{\partial \mathbf{w}} =$

$\frac{\partial \mathcal{L}}{\partial \mathbf{w}} =$

5. Napište nám cokoliv co se vám zatím libilo/nelíbilo a jak bychom to měli změnit. Pokud máte stále dost času, nakreslete nám obrázek na téma **"Učení robotů a Halloween"**. Vybraná díla budou zveřejněna na stránkách předmětu, nejkreativnější výtvor bude **odměněn lahváčem**.

Nelíbilo se mi, že všechny ty příklady jsme neprobírali na cvičeních. Líbí se mi systém přípravy na cvičení předmětu OPT, dostáváme každý týden pár úloh a máme je vyřešit ke příštímu cvičení. Na cvičení probíráme jejich řešení. Podle mě bylo by dobrý implementovat podobný systém v tomto předmětu

Obyčejný kvíz na přednášce UROB:



a) Evil Morty
b) Jack
c) Vedoucí katedry elektrotechniky
d) Andrej Babiš
e) Nevím