גרדיאנט שכבת הקוונטיזציה

13 ביוני 2019

נסתכל על שכבת קוונטיזציה עם N קוונטייזרים, M מילות קוד כל אחד. נסמן את וקטור הכניסה לשכבה בתור X, וקטור היציאה בתור Z ואת וקטור המוצא של כל הרשת בתור L. X ו־Z הינם ממימד N ואילו L, במקרה כללי היה ממימד כלשהו P אבל מכיוון שבמטלב מוצא הרשת חייב להיות סקלרי (ה־loss) L הינו סקלר.

- מוצא השכבה **בשלב האימון** נתון ע"י:

$$Z = \begin{pmatrix} Z_1 & \cdots & Z_l & \cdots & Z_N \end{pmatrix}$$

כאשר:

$$\boxed{Z_i = \sum_{j=1}^{M} a_j \tanh\left[c_j \left(X_i - b_j\right)\right]}$$

- גרדיאנט מוצא השכבה לפי מבוא השכבה הינו:

$$\frac{\partial Z}{\partial X} = \begin{pmatrix} \frac{\partial Z_1}{\partial X_1} & \cdots & \frac{\partial Z_1}{\partial X_i} & \cdots & \frac{\partial Z_1}{\partial X_N} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \frac{\partial Z_i}{\partial X_1} & \cdots & \frac{\partial Z_i}{\partial X_i} & \cdots & \frac{\partial Z_i}{\partial X_N} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \frac{\partial Z_N}{\partial X_1} & \cdots & \frac{\partial Z_N}{\partial X_i} & \cdots & \frac{\partial Z_N}{\partial X_N} \end{pmatrix} = \begin{pmatrix} \frac{\partial Z_1}{\partial X_1} & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{\partial Z_i}{\partial X_i} & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & \frac{\partial Z_N}{\partial X_N} \end{pmatrix}$$

שכן הקוונטייזרים נפרדים והנגזרת של מוצא מסויים לפי מבוא אחר מתאפסת:

$$\frac{\partial Z_l}{\partial X_i} = 0 \quad , \quad \forall l \neq i$$

עבור נגזרות שאינן מתאפסות:

$$\frac{\partial Z_i}{\partial X_i} = \frac{\partial}{\partial X_i} \sum_{j=1}^{M} a_j \tanh\left[c_j \left(X_i - b_j\right)\right] = \sum_{j=1}^{M} a_j \frac{\partial}{\partial X_i} \tanh\left[c_j \left(X_i - b_j\right)\right] = \sum_{j=1}^{M} a_j c_j \left\{1 - \tanh^2\left[c_j \left(X_i - b_j\right)\right]\right\}$$

- נגזרות של Z לפי פרמטרי השכבה:
  - נגזרת לפי $b = \begin{pmatrix} b_1 & \cdots & b_i & \cdots & b_M \end{pmatrix}$:

$$\frac{\partial Z}{\partial b} = \begin{pmatrix} \frac{\partial Z_1}{\partial b_1} & \cdots & \frac{\partial Z_1}{\partial b_i} & \cdots & \frac{\partial Z_1}{\partial b_M} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \frac{\partial Z_l}{\partial b_1} & \cdots & \frac{\partial Z_l}{\partial b_i} & \cdots & \frac{\partial Z_l}{\partial b_M} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \frac{\partial Z_N}{\partial b_1} & \cdots & \frac{\partial Z_N}{\partial b_i} & \cdots & \frac{\partial Z_N}{\partial b_M} \end{pmatrix}$$

$$\frac{\partial Z_l}{\partial b_i} = \frac{\partial}{\partial b_i} \sum_{j=1}^{M} a_j \tanh\left[c_j \left(X_i - b_j\right)\right] = a_i c_j \left\{\tanh^2\left[c_j \left(X_i - b_j\right)\right] - 1\right\}$$

- נחלק את גרדיאנט מוצא הרשת L לפי הכניסות לשכבת הקוונטיזציה למקרה בו מוצא הרשת הינו ממימד P ולמקרה בו הינו סקלר:
  - **<u>וקטור:</u>** עבור מוצא רשת וקטורי נקבל:
    - גרדיאנט מוצא השכבה, L, לפי הכניסות לשכבה עבור מוצא ממימד P:

$$\frac{\partial L}{\partial X}=\begin{pmatrix} \frac{\partial L_1}{\partial X_1} & \cdots & \frac{\partial L_1}{\partial X_i} & \cdots & \frac{\partial L_1}{\partial X_N} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \frac{\partial L_k}{\partial X_1} & \cdots & \frac{\partial L_k}{\partial X_i} & \cdots & \frac{\partial L_k}{\partial X_N} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \frac{\partial L_P}{\partial X_1} & \cdots & \frac{\partial L_P}{\partial X_i} & \cdots & \frac{\partial L_P}{\partial X_N} \end{pmatrix}$$

כאשר איבר כללי ביעקוביאן הינו:

$$\frac{\partial L_k}{\partial X_i}=\sum_{l=1}^{N}\frac{\partial L_k}{\partial Z_l}\frac{\partial Z_l}{\partial X_i}=\frac{\partial L_k}{\partial Z_i}\frac{\partial Z_i}{\partial X_i}=\frac{\partial L_k}{\partial Z_i}\sum_{j=1}^{M}a_jc_j\left\{1-\tanh^2\left[c_j\left(X_i-b_j\right)\right]\right\}$$

    - גרדיאנט מוצא השכבה, L, לפי פרמטרי השכבה:
      - נגזרת לפי b:

$$\frac{\partial L}{\partial b}=\begin{pmatrix} \frac{\partial L_1}{\partial b_1} & \cdots & \frac{\partial L_1}{\partial b_i} & \cdots & \frac{\partial L_1}{\partial b_M} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \frac{\partial L_k}{\partial b_1} & \cdots & \frac{\partial L_k}{\partial b_i} & \cdots & \frac{\partial L_k}{\partial b_M} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \frac{\partial L_P}{\partial b_1} & \cdots & \frac{\partial L_P}{\partial b_i} & \cdots & \frac{\partial L_P}{\partial b_M} \end{pmatrix}$$

ואיבר כללי הינו:

$$\frac{\partial L_k}{\partial b_i}=\sum_{l=1}^{N}\frac{\partial L_k}{\partial Z_l}\frac{\partial Z_l}{\partial b_i}=a_ic_i\sum_{l=1}^{N}\frac{\partial L_k}{\partial Z_l}\left\{\tanh^2\left[c_j\left(X_l-b_i\right)\right]-1\right\}$$

  - **<u>סקלר:</u>** נפשט את הביטויים עבור מוצא וקטורי של הרשת ונקבל:
    - גרדיאנט לפי הכניסות לשכבה:

$$\frac{\partial L}{\partial X}=\begin{pmatrix} \frac{\partial L}{\partial X_1} & \cdots & \frac{\partial L}{\partial X_i} & \cdots & \frac{\partial L}{\partial X_N} \end{pmatrix}$$

כאשר איבר כללי הינו:

$$\boxed{\frac{\partial L}{\partial X_i}}=\sum_{l=1}^{N}\frac{\partial L}{\partial Z_l}\frac{\partial Z_l}{\partial X_i}=\frac{\partial L}{\partial Z_i}\frac{\partial Z_i}{\partial X_i}\boxed{=\frac{\partial L}{\partial Z_i}\sum_{j=1}^{M}a_jc_j\left\{1-\tanh^2\left[c_j\left(X_i-b_j\right)\right]\right\}}$$

  - גרדיאנט מוצא השכבה, L, לפי פרמטרי השכבה:
    - נגזרת לפי b:

$$\frac{\partial L}{\partial b}=\begin{pmatrix} \frac{\partial L}{\partial b_1} & \cdots & \frac{\partial L}{\partial b_i} & \cdots & \frac{\partial L}{\partial b_M} \end{pmatrix}$$

ואיבר כללי הינו:

$$\boxed{\frac{\partial L}{\partial b_i}}=\sum_{l=1}^{N}\frac{\partial L}{\partial Z_l}\frac{\partial Z_l}{\partial b_i}\boxed{=a_ic_i\sum_{l=1}^{N}\frac{\partial L}{\partial Z_l}\left\{\tanh^2\left[c_j\left(X_l-b_i\right)\right]-1\right\}}$$