Weekly Report (4)

> ----By Chen Junlin at UESTC
>
> chenjunlin@std.uestc.edu.cn

>
> **Week:** 2019-2020-1学期 第10周
>
> **Time:** 2019/11/4 ~ 2019/11/10

本周完成内容

由于期中考试周，未能有太多进展.

对于作业1中的SVM方法，在写到推导SVM损失函数的 Gradient 时，遇到了困难（主要是对涉及到向量和矩阵的求导、求偏导、求梯度等数学工具不够熟悉，以及将推导出的梯度公式在代码中实现），于是就卡住了。。

以下是推导的一些过程。

X_train $(49000, 3072) + \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}_{49000:1}$ $(49000, 3073)$

X_val $(1000, 3073)$

X_test $(1000, 3073)$

X_dev $(500, 3073)$ (正确/训练) y_dev $(500,)$

scores $= \underbrace{X[i]}_{\text{dev } 1\times3073} \cdot W_{3073\times10}$ × = [—] ↑ 1×10

correct_ = scores[y[i]]

margin = scores[j] - correct_ + 1

loss += $\frac{\text{margin}}{\text{num\_train}}$

loss + $\lambda R(W)$

$$R(W) = \sum_k \sum_l W_{k,l}^2$$

SUM loss → measure unhappiness

i-th example:

pixels: x_i [3072 × 1] + bias 1 = [3073 × 1] ← column vector

label: y_i (correct class)

scores: $f(x_i, W)_j = s_j$ of j-th class 10×3073 3073×1

$$L_i = \sum_{j \neq y_i} \max(0, s_j - s_{y_i} + \Delta) \quad (s_j = W x_j)$$

↑ for $j = y_i$, $s_j - s_{y_i} = 0$

$$= \sum_{j \neq y_i} \max(0, w_j^T x_i - w_{y_i}^T x_i + \Delta)$$

↑ the j-th row reshaped as column vector.

Gradient of loss: w.r.t. w_{y_i}

$\boxed{\nabla f(p) = \left[\frac{\partial f}{\partial x_1}(p) \right]}$

$$\frac{\partial f}{\partial x_n}(p)$$

$$\nabla_{w_{y_i}} L_i = \sum_{j \neq y_i} \nabla_{w_{y_i}} \max(0, w_j^T x_i - w_{y_i}^T x_i + \Delta)$$

$$= \sum_{j \neq y_i} \begin{cases} \nabla_{w_{y_i}} 0 = 0, & (\ ) < 0 \\ \nabla_{w_{y_i}} (\ ) = -x_i, & (\ ) > 0 \end{cases}$$

$$\boxed{\frac{\partial \vec{x}^T \vec{a}}{\partial \vec{x}} = \frac{\partial \vec{a}^T \vec{x}}{\partial \vec{x}} = \vec{a}}$$

$$= \sum_{j \neq y_i} I(\ ) \cdot (-x_i)$$

$$\boxed{I(x) = \begin{cases} 1, & x = true \\ 0, & x = false \end{cases}}$$

$$= -\left(\sum_{j \neq y_i} I\left(w_j^T x_i - w_{y_i}^T x_i + \Delta > 0\right) \right) x_i$$

Vector Calculus: $\vec{y} = [y_1\ y_2 \cdots y_m]^T$

$$\frac{\partial \vec{y}}{\partial x} = \begin{bmatrix} \frac{\partial y_1}{\partial x} \\ \vdots \\ \frac{\partial y_m}{\partial x} \end{bmatrix} \qquad \frac{\partial y}{\partial \vec{x}} = \begin{bmatrix} \frac{\partial y}{\partial x_1} & \frac{\partial y}{\partial x_2} & \cdots & \frac{\partial y}{\partial x_n} \end{bmatrix}$$

$$\vec{x} = [x_1\ x_2 \cdots x_n]^T$$

$$\therefore \nabla_{w_{y_i}} (-w_{y_i}^T x_i)$$

$$= \begin{bmatrix} \frac{\partial}{\partial w_{y_i}(1)} \left(-[w_{y_i}(1) \cdots] \begin{bmatrix} x_i(1) \\ \vdots \\ \vdots \end{bmatrix}\right) \\ \cdots \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\partial}{\partial w_{y_i}(1)} \left[-(w_{y_i}(1)\, x_i(1) + \cdots)\right] \\ \cdots \end{bmatrix}$$

$$= \begin{bmatrix} -x_i(1) \\ \vdots \end{bmatrix}$$

$$= -x_i$$

困难和挑战

如上，在数学推导方面遇到困难，以及将涉及到矩阵的数学公式用代码实现时，感觉无法绕过弯来。。

下周计划

下周五还有一门期中考试，忙过这段时间后会好好补起来缺失的内容。