

COMP90051

Statistical Machine Learning

Workshop Week 5

Xudong Han

https://github.com/HanXudong/COMP90051_Workshops

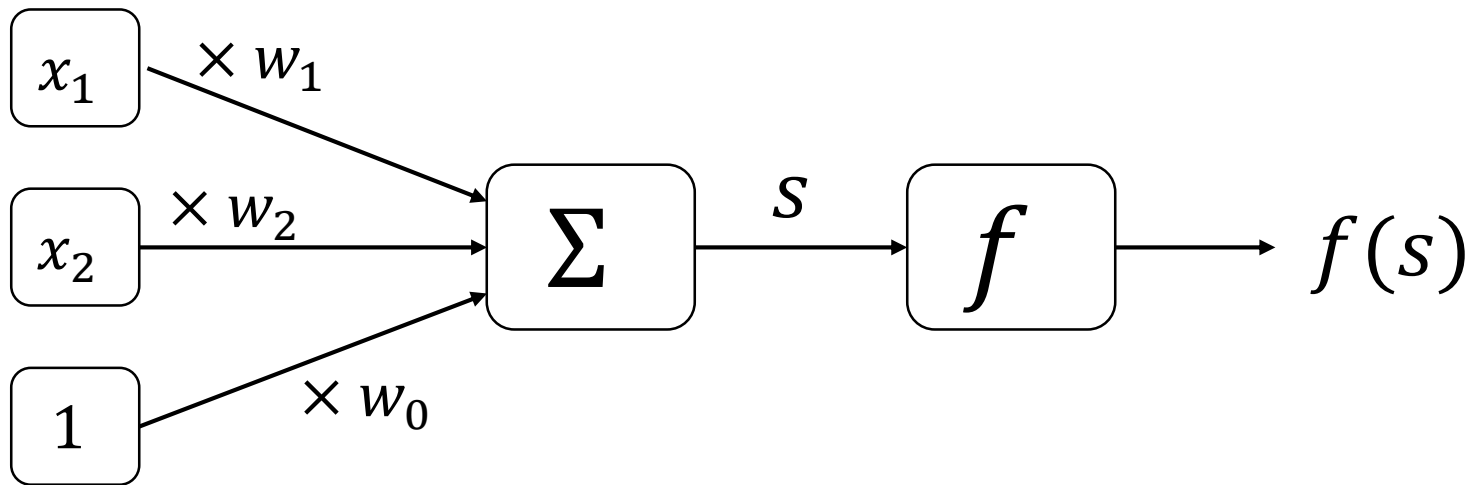
Your tutor

- Xudong Han
- xudong.han@unimelb.edu.au
- Slides
https://github.com/HanXudong/COMP90051_Worksheets

Logistic Regression = Linear Regression + Logistic function

- $\theta = p(y = 1|\vec{x}) = \sigma(\vec{w}^T \vec{X} + b)$
- $\vec{w}^T = (w_1 \quad w_2 \quad \dots \quad w_p)$
- $\vec{X}^T = (x_1 \quad x_2 \quad \dots \quad x_p)$
- $b = w_0$
- Loss function: Cross-Entropy

Definition of the perceptron



- $$f(s) = \begin{cases} 1 & \text{if } s \geq 0, \\ 0 & \text{otherwise} \end{cases}$$

Perceptron training algorithm

PERCEPTRON(\mathbf{w}_0)

```
1   $\mathbf{w}_1 \leftarrow \mathbf{w}_0$        $\triangleright$  typically  $\mathbf{w}_0 = \mathbf{0}$ 
2  for  $t \leftarrow 1$  to  $T$  do
3      RECEIVE( $\mathbf{x}_t$ )
4       $\hat{y}_t \leftarrow \text{sgn}(\mathbf{w}_t \cdot \mathbf{x}_t)$ 
5      RECEIVE( $y_t$ )
6      if  $(\hat{y}_t \neq y_t)$  then
7           $\mathbf{w}_{t+1} \leftarrow \mathbf{w}_t + y_t \mathbf{x}_t$      $\triangleright$  more generally  $\eta y_t \mathbf{x}_t, \eta > 0$ .
8      else  $\mathbf{w}_{t+1} \leftarrow \mathbf{w}_t$ 
9  return  $\mathbf{w}_{T+1}$ 
```

Perceptron training algorithm

- The above training procedure is equivalent to performing sequential gradient descent on the following objective function:

$$F(\mathbf{w}) = \frac{1}{T} \sum_{t=1}^T \max(0, -y_t(\mathbf{w} \cdot \mathbf{x}_t))$$

$$\mathbf{w}_{t+1} \leftarrow \mathbf{w}_t - \eta \nabla_{\mathbf{w}} F(\mathbf{w})$$

Evaluation

- proportion of misclassified instances (error rate)