Overfitting

Choosing activation function

Activation Functions

Sigmoid

$$Sigmoid(x) = \frac{1}{1+\exp(-x)}$$

$$\frac{d}{dx}\text{sigmoid}(x) = \frac{\exp(-x)}{(1+\exp(-x))^2} = sigmoid(x)\big(1 - sigmoid(x)\big)$$

ReLU

$$ReLU(x) = \max(x, 0)$$

$$pReLU(x) = \max(0, x) + \alpha\min(0, x)$$

Softmax

$$Softmax(x_i) = \frac{e^{x_i}}{\sum_{k=1}^{K} e^{x_k}}$$

Subtopic

Hidden Layers

❖Hidden layers
- ➢By default, ReLU

❖Output layer
- ➢Depends on your task
- ➢Regression: Sigmoid
- ➢Binary classification: Sigmoid
- ➢Multi-class classificaton: Softmax

Multi-class classification

❖One-hot encoding
- ➢Dog: [1, 0, 0]
- ➢Cat: [0, 1, 0]
- ➢Bird: [0, 0, 1]

❖Cross-entropy loss
- ➢$CE(Y, d) = -\sum_i^k d_i log y_i$
- ➢Y: The vector after Softmax
- ➢d: The one-hot encoded label vector

$$Softmax(x_i) = \frac{e^{x_i}}{\sum_{k=1}^{K} e^{x_k}}$$

Overfitting

What is overfitting

Number of Parameters

Too Complex -- Overfitting

Too Simple -- Underfitting

Evaluation

❖Train-validation-test split
- ➢Train: 70%
- ➢Validation: 15%
- ➢Test: 15%

❖Cross-validation
- ➢5-fold validation
- ➢Leave-one-out
- ➢Reliable evaluation
- ➢Expensive

What to evaluate

Loss Function

Performance

Complexity

Data
Too little, not reflect the true distribution

Model
Too large, too many useless parameters

Connectivity
Too strong, co-adaptation

Parameter value range
Too large, model too flexible

Training time
Too long, tend to overfitting