220110724 陈彦至 Homework 5

Derivation of Bias-Variance Decomposition of Mean Squared Error

1. **Bias-Variance Decomposition**:

   - Consider a true model $f(x)$ and an estimated model $\hat{f}(x)$. The expected prediction error at a point x is given by:

     $$E[(y - \hat{f}(x))^2]$$

     where $y = f(x) + \epsilon$ and ϵ is noise with mean zero and variance σ^2.

   - Expanding the squared error:

     $$E[(f(x) + \epsilon - \hat{f}(x))^2] = E[(f(x) - \hat{f}(x))^2] + 2E[(f(x) - \hat{f}(x))\epsilon] + E[\epsilon^2]$$

   - Since ϵ is independent and has zero mean, the cross term vanishes:

     $$E[(f(x) - \hat{f}(x))^2] + \sigma^2$$

   - Decompose the remaining term by adding and subtracting $E[\hat{f}(x)]$:

     $$E[(f(x) - E[\hat{f}(x)] + E[\hat{f}(x)] - \hat{f}(x))^2]$$

   - Expanding and simplifying:

     $$(f(x) - E[\hat{f}(x)])^2 + E[(E[\hat{f}(x)] - \hat{f}(x))^2]$$

   - This results in:

     $$Bias^2(\hat{f}(x)) + Var(\hat{f}(x)) + \sigma^2$$

Summary

$$E[(y - \hat{f}(x))^2] = Bias^2(\hat{f}(x)) + Var(\hat{f}(x)) + \sigma^2$$

Derivation of Back Propagation for Fully Connected 3-Layered Network

2. **Backpropagation for a 3-Layered Network**:

   - **Forward Pass**:
     - Input to hidden: $z^1 = W^1 x + b^1$
     - Hidden activation: $h = \sigma(z^1)$

    - Hidden to output: $z^2 = W^2 h + b^2$
    - Output activation: $o = \phi(z^2)$
    - Loss: $L = \frac{1}{2} \sum_{i=1}^{k} (o_i - y_i)^2$
- **Backward Pass**:
    - Output layer error: $\delta^2 = (o - y) \odot \phi'(z^2)$
    - Hidden layer error: $\delta^1 = ((W^2)^T \delta^2) \odot \sigma'(z^1)$
    - Gradients:
        - Weight matrix W^2: $\frac{\partial L}{\partial W^2} = \delta^2 h^T$
        - Bias vector b^2: $\frac{\partial L}{\partial b^2} = \delta^2$
        - Weight matrix W^1: $\frac{\partial L}{\partial W^1} = \delta^1 x^T$
        - Bias vector b^1: $\frac{\partial L}{\partial b^1} = \delta^1$

Summary

$$\delta^2 = (o - y) \odot \phi'(z^2)$$
$$\delta^1 = ((W^2)^T \delta^2) \odot \sigma'(z^1)$$
$$\frac{\partial L}{\partial W^2} = \delta^2 h^T$$
$$\frac{\partial L}{\partial b^2} = \delta^2$$
$$\frac{\partial L}{\partial W^1} = \delta^1 x^T$$
$$\frac{\partial L}{\partial b^1} = \delta^1$$