Regression

- Linear activation function in the output layer ensures an arbitrary output range

[Diagram showing input, output, and target layers with linear activation functions]
Classification

- Sigmoid activation function in the output layer ensures outputs between 0 and 1
Classification test example (after learning)

Input

$\mathbf{x}_N$

$\mathbf{x}_1$

$\mathbf{x}_0$

Output

Softmax

$z_K$

$z_1$

Activation functions

Softmax

K categories

0.9 0.54

“car”

0.03 0.225

“motorbike”

0.07 0.235

“cow”
NN and overfitting

- Complexity grows with number of hidden neurons
- Overfitting will occur (Test error >> Training error)
- Three methods to prevent overfitting:
  - Cross-validation and model selection
  - Weight decay
  - Early stopping