

Sources of error

2. (10 points) Recall that *structural* error arises when the hypothesis class cannot represent a hypothesis that performs well on the test data and *estimation* error arises when the parameters of a hypotheses cannot be estimated well based on the training data.

Following is a collection of potential cures for a situation in which your learning algorithm generates a hypothesis with high test error.

For each one, indicate whether it **can reduce** structural error, estimation error, neither, or both.

- (a) Penalize $\|\theta\|^2$ during training
☐ structural error ☒ **estimation error** ☐ both ☐ neither
- (b) Penalize $\|\theta\|^2$ during testing
☐ structural error ☐ estimation error ☐ both ☒ **neither**
- (c) Increase the amount of training data
☐ structural error ☒ **estimation error** ☐ both ☐ neither
- (d) Increase the order of a fixed polynomial basis
☒ **structural error** ☐ estimation error ☐ both ☐ neither
- (e) Decrease the order of a fixed polynomial basis
☐ structural error ☒ **estimation error** ☐ both ☐ neither
- (f) Add more layers with linear activation functions to your neural network
☐ structural error ☐ estimation error ☐ both ☒ **neither**
- (g) Add more layers with non-linear activation functions to your neural network
☒ **structural error** ☐ estimation error ☐ both ☐ neither
- (h) Stop training before training error reaches 0
☐ structural error ☒ **estimation error** ☐ both ☐ neither

Solution: Recall that both structural and estimation error are defined in terms of error on the test data. Structural error is inherent to the hypothesis class being insufficiently rich to represent the data, while estimation error occurs when training does not achieve good generalizable parameters.

a) Penalizing $\|\theta\|^2$ during training can reduce estimation error because it can prevent overfitting on training data. Therefore, it can improve generalization to test data.

b) Penalizing $\|\theta\|^2$ during testing can have no effect because θ is not updated during testing. The additional penalty from the regularization term of the objective would be the same for every test point.

c) Increasing the amount of training data can reduce estimation error because more updates to the parameters of the hypothesis class during training can result in better predictions in testing, in the case of no overfitting.

d) Increasing the order of a fixed polynomial basis is an example of increasing the complexity of the hypothesis class. If we begin with an overly simple model (high structural

Name: _____

error) on data generated from a complicated distribution, making the model more complex could achieve better performance on test data.

e) Decreasing the order of a fixed polynomial basis is an example of decreasing the complexity of the hypothesis class. We can potentially combat overfitting by reducing model complexity, so decreasing polynomial order could reduce estimation error.

f) Adding layers with linear activations will not affect model complexity.

g) Adding layers with non-linear activations is another example of increasing the complexity of the hypothesis class, which can result in better predictions in testing.

h) Stopping training early, before training error reaches zero, is one way to prevent overfitting. Therefore it can reduce estimation error.

For each of the following situations, indicate whether the **poor performance is due to** high structural error, high estimation error, neither, or both.

- (i) Neural network has very low training error but high testing error.
☐ structural error ☒ **estimation error** ☐ both ☐ neither
- (j) Neural network training error is persistently high, as is test error.
☒ **structural error** ☐ estimation error ☐ both ☐ neither

Solution: i) A neural network with low training error but high testing error has overfit, so it has high estimation error. We could potentially combat this by regularizing, choosing a simpler hypothesis class, or by stopping training early.

j) If neural network training error and test error are both persistently high, we are probably using the wrong type of model. Therefore, structural error is high and we should consider other hypothesis classes.