Week6 Advice for Applying Machine Learning

Deciding What to Try Next

Which of the following statements about diagnostics are true? Check all that apply.

- [ ] It's hard to tell what will work to improve a learning algorithm, so the best approach is to go with gut feeling and just see what works.

未选择的是正确的

- [x] Diagnostics can give guidance as to what might be more fruitful things to try to improve a learning algorithm.

正确

- [x] Diagnostics can be time-consuming to implement and try, but they can still be a very good use of your time.

正确

- [x] A diagnostic can sometimes rule out certain courses of action (changes to your learning algorithm) as being unlikely to improve its performance significantly.

正确

Evaluating a Hypothesis

Suppose an implementation of linear regression (without regularization) is badly overfitting the training set. In this case, we would expect:

- (●) The training error $J(\theta)$ to be **low** and the test error $J_{\text{test}}(\theta)$ to be **high**

正确

- ( ) The training error $J(\theta)$ to be **low** and the test error $J_{\text{test}}(\theta)$ to be **low**
- ( ) The training error $J(\theta)$ to be **high** and the test error $J_{\text{test}}(\theta)$ to be **low**
- ( ) The training error $J(\theta)$ to be **high** and the test error $J_{\text{test}}(\theta)$ to be **high**

过拟合：训练集代价低，测试集代价高

Model Selection and Train/Validation/Test Sets

Consider the model selection procedure where we choose the degree of polynomial using a cross validation set. For the final model (with parameters θ), we might generally expect $J_{\mathrm{CV}}(\theta)$ To be lower than $J_{\mathrm{test}}(\theta)$ because:

- ◉ An extra parameter (d, the degree of the polynomial) has been fit to the cross validation set.

正确

- ○ An extra parameter (d, the degree of the polynomial) has been fit to the test set.
- ○ The cross validation set is usually smaller than the test set.
- ○ The cross validation set is usually larger than the test set.

多项式维度d是根据CV选择的，选取的是在CV上代价最低的d作为最终模型，所以通常CV代价会低于测试集代价

Diagnosing Bias vs. Variance

Suppose you have a classification problem. The (misclassification) error is defined as $\frac{1}{m}\sum_{i=1}^{m}\mathrm{err}(h_\theta(x^{(i)}),y^{(i)})$, and the cross validation (misclassification) error is similarly defined, using the cross validation examples $(x_{\mathrm{cv}}^{(1)},y_{\mathrm{cv}}^{(1)}),\ldots,(x_{\mathrm{cv}}^{(m_{\mathrm{cv}})},y_{\mathrm{cv}}^{(m_{\mathrm{cv}})})$. Suppose your training error is 0.10, and your cross validation error is 0.30. What problem is the algorithm most likely to be suffering from?

- ○ High bias (overfitting)
- ○ High bias (underfitting)
- ◉ High variance (overfitting)

正确

- ○ High variance (underfitting)

高偏差/欠拟合：训练集代价与CV代价均较高且接近

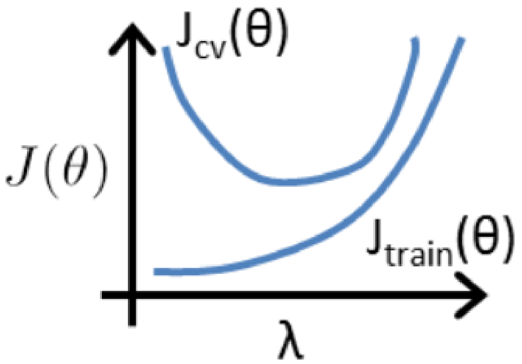
高方差/过拟合：训练集代价较低而CV代价较高，二者之间的差异随训练样本增多逐渐减小但仍存在

Regularization and Bias/Variance

Consider regularized logistic regression. Let

- $J(\theta) = \frac{1}{2m}[\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2 + \lambda \sum_{j=2}^{n} \theta_j^2]$
- $J_{\text{train}}(\theta) = \frac{1}{2m_{\text{train}}}[\sum_{i=1}^{m_{\text{train}}}(h_\theta(x_{\text{train}}^{(i)}) - y_{\text{train}}^{(i)})^2]$
- $J_{\text{CV}}(\theta) = \frac{1}{2m_{\text{CV}}}[\sum_{i=1}^{m_{\text{CV}}}(h_\theta(x_{\text{CV}}^{(i)}) - y_{\text{CV}}^{(i)})^2]$

Suppose you plot J_{train} and J_{CV} as a function of the regularization parameter λ. which of the following plots do you expect to get?



正确

正则化参数大：高偏差，欠拟合

正则化参数小：高方差，过拟合

Learning Curves

In which of the following circumstances is getting more training data likely to significantly help a learning algorithm's performance?

- [ ] Algorithm is suffering from high bias.

未选择的是正确的

- [x] Algorithm is suffering from high variance.

正确

- [x] $J_{\text{CV}}(\theta)$ (cross validation error) is much larger than $J_{\text{train}}(\theta)$ (training error).

正确

- [ ] $J_{\text{CV}}(\theta)$ (cross validation error) is about the same as $J_{\text{train}}(\theta)$ (training error).

未选择的是正确的

高方差下增大样本集可以提高性能

Deciding What to Do Next Revisited

Suppose you fit a neural network with one hidden layer to a training set. You find that the cross validation error $J_{\text{CV}}(\theta)$ is much larger than the training error $J_{\text{train}}(\theta)$. Is increasing the number of hidden units likely to help?

- ( ) Yes, because this increases the number of parameters and lets the network represent more complex functions.
- ( ) Yes, because it is currently suffering from high bias.
- ( ) No, because it is currently suffering from high bias, so adding hidden units is unlikely to help.
- (•) No, because it is currently suffering from high variance, so adding hidden units is unlikely to help.

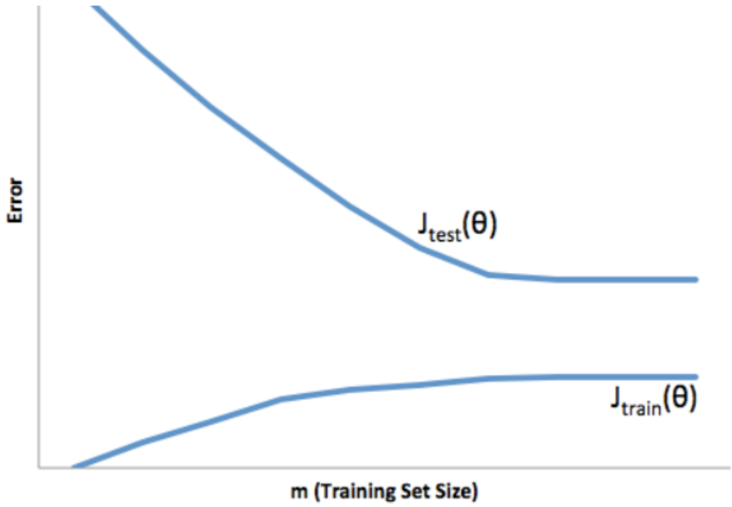
正确

CV代价远高于训练集代价——过拟合，应减少特征量而非增加

Advice for Applying Machine Learning

1
分数

1. You train a learning algorithm, and find that it has unacceptably high error on the test set. You plot the learning curve, and obtain the figure below. Is the algorithm suffering from high bias, high variance, or neither?



- ( ) High bias
- (•) High variance
- ( ) Neither

1
分数

2. Suppose you have implemented regularized logistic regression to classify what object is in an image (i.e., to do object recognition). However, when you test your hypothesis on a new set of images, you find that it makes unacceptably large errors with its predictions on the new images. However, your hypothesis performs **well** (has low error) on the training set. Which of the following are promising steps to take? Check all that apply.

- [ ] Try evaluating the hypothesis on a cross validation set rather than the test set.
- [x] Try increasing the regularization parameter λ.
- [ ] Try decreasing the regularization parameter λ.
- [x] Try using a smaller set of features.

1
分数

3. Suppose you have implemented regularized logistic regression to predict what items customers will purchase on a web shopping site. However, when you test your hypothesis on a new set of customers, you find that it makes unacceptably large errors in its predictions. Furthermore, the hypothesis performs **poorly** on the training set. Which of the following might be promising steps to take? Check all that apply.

- [ ] Use fewer training examples.
- [x] Try decreasing the regularization parameter λ.
- [ ] Try evaluating the hypothesis on a cross validation set rather than the test set.
- [x] Try adding polynomial features.

1
分数

4. Which of the following statements are true? Check all that apply.

- [ ] It is okay to use data from the test set to choose the regularization parameter λ, but not the model parameters (θ).
- [x] A typical split of a dataset into training, validation and test sets might be 60% training set, 20% validation set, and 20% test set.
- [ ] Suppose you are training a logistic regression classifier using polynomial features and want to select what degree polynomial (denoted d in the lecture videos) to use. After training the classifier on the entire training set, you decide to use a subset of the training examples as a validation set. This will work just as well as having a validation set that is separate (disjoint) from the training set.
- [x] Suppose you are using linear regression to predict housing prices, and your dataset comes sorted in order of increasing sizes of houses. It is then important to randomly shuffle the dataset before splitting it into training, validation and test sets, so that we don't have all the smallest houses going into the training set, and all the largest houses going into the test set.

1
分数

5. Which of the following statements are true? Check all that apply.

- [ ] We always prefer models with high variance (over those with high bias) as they will able to better fit the training set.
- [x] If a learning algorithm is suffering from high variance, adding more training examples is likely to improve the test error.
- [x] If a learning algorithm is suffering from high bias, only adding more training examples may **not** improve the test error significantly.
- [x] When debugging learning algorithms, it is useful to plot a learning curve to understand if there is a high bias or high variance problem.

我们通常选择较复杂的神经网络，采取正则化方法克服过拟合问题，而非选择欠拟合问题严重的简单神经网络