Prior Solutions:

1. Winning Solution
   1. Features Selection
      1. **Normalized all features to a zero mean with a standard deviation of 1.**
      2. **Dropped \*-phi features, because they increased overfitting.**
      3. Features with long-tails were log transformed.
   2. Modeling Techniques and Training
      1. Neural networks – averaged the probabilities of an example being S, top 15.2% of the examples. The cutoff was hand selected based on AMS vs cutoff curve in cross-validation.
         1. 3 hidden layers, 600 neurons
         2. 2 nueron softmax output layer, S and B
         3. Cross entropy loss and backpropagation
         4. Tip: Use 8-16 neural networks instead of 70
         5. Tip2 : switch to single-precision floating point format and use 1 neural network
         6. Max depth of 9, 3000 steps.
   3. Interview
      1. Feature engineering is not always key. Locally evaluating solutions is. Local cross-validation
      2. Need clear explanation of why AMS improved overall.
2. Second Place Solution
   1. Features Selection
      1. **Removed the phi features.**
   2. Modeling and Training
      1. Used Boosted decision trees.
      2. 7 fold cross validation; train a model on complete training set
      3. Each script on a machine with 8+ cores and took 2 days.
      4. Trained 8 models in parallel
   3. Feature Engineering
      1. Transforming the eta and phi features.
         1. <https://www.kaggle.com/c/higgs-boson/forums/t/9576/reducing-the-feature-space>
   4. Prediction target
      1. <https://github.com/TimSalimans/HiggsML>
   5. Regularized Greedy Forest -> similar to GBM or XGBoost (Python)
3. Third Place Solution
   1. Features Selection/Extraction
      1. **Normalized them all to have zero mean, 1 standard deviation.**
      2. Cake features did not help.
   2. Missing data imputation
      1. Regression imputation worsened results.
      2. Simple imputation rules (min, median, mean, max) used in equal proportions.
   3. Modeling Techniques and Training
      1. 108 neural networks.
      2. Averaged the predictions. 0 to 1. Signal if > .5575. Based on PL and CV.
      3. Ensemble:
         1. 30x50x1
         2. 30x50x25x1
         3. 30x50x50x25x1
         4. <https://www.kaggle.com/c/higgs-boson/forums/t/10481/third-place-model-documentation/55390#post55390>
   4. Things that did not work:
      1. Boosting
      2. Bagging
      3. Kmeans
      4. Stacking
      5. Attribute selection with GA