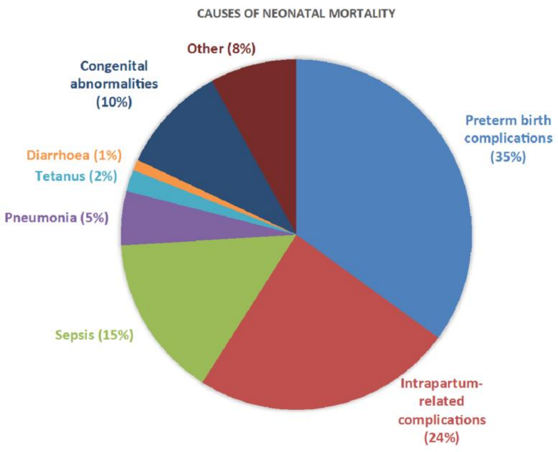
Classification of Fetal State by Cardiotocography

BMES-547: Machine Learning

Ellen Garven and Vivian Wu

Biomedical Problem [1,2]

- Direct comparisons of neonatal morbidity and mortality are difficult because of varying definitions and misclassifications.
- Under-five mortality has decreased:
  - 1990: 90 deaths per 1000 live births
  - 2013: 46 deaths per 1000 live births
- Fetal distress is one of the main factors of neonatal mortality.



Narayanan, Indira & Vivio, Donna. (2016).

Cardiotocography (CTG) [2]

- Non-invasive technique for monitoring fetal conditions in the antepartum period.
  - Two transducers simultaneously measure fetal heart rate (FHR) and uterine contractions.
- Conditions such as hypoxia, academia, drug induction produce noticeable variations of FHR.
- FHR is an important index to identify occurrences of fetal distress.

The Dataset [3,4]

- 2126 fetal CTGs classified by 3 expert obstetricians according to fetal state class code to indicate existence of fetal distress.
  - 21 given attributes
  - 1655 **normal**, 295 **suspicious**, 176 **pathologic**
- SisPorto automatically processed CTGs and analyzed ante- and intrapartum tracings.
  - Indicated by *

LB*—FHR baseline (beats per minute)
AC*—# of accelerations per second
FM*—# of fetal movements per second
UC*—# of uterine contractions per second
DL—# of light decelerations per second
DS—# of severe decelerations per second
DP—# of prolonged decelerations per second
ASTV*—% of time with abnormal short term variability
MSTV*—mean value of short term variability
ALTV*—% of time with abnormal long term variability
MLTV*—mean value of long term variability
Width—width of FHR histogram
Min—minimum of FHR histogram
Max—Maximum of FHR histogram
Nmax—# of histogram peaks
Nzeros—# of histogram zeros
Mode—histogram mode
Mean—histogram mean
Median—histogram median
Variance—histogram variance
Tendency—histogram tendency
NSP—fetal state class code

Pre-processing

- Extracted predictor data into X (21 attributes)
- Extracted last column as class labels into T (fetal state class code)

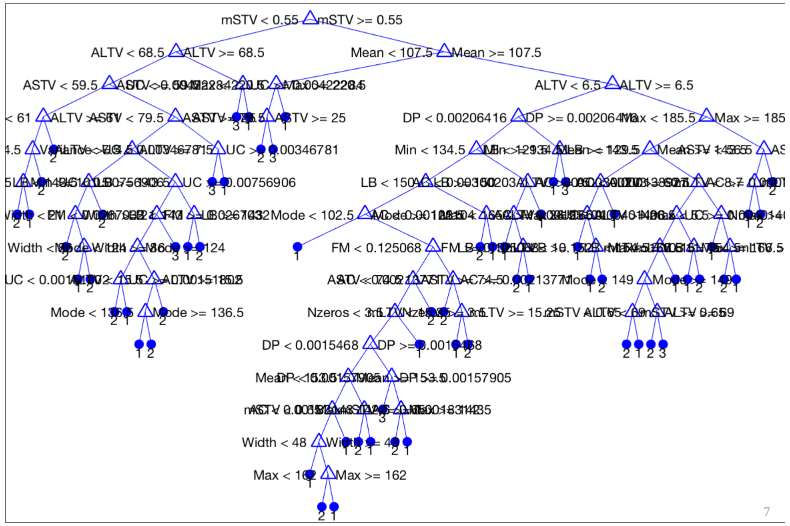
Methods

- Decision Tree for Classification, Visualization, Feature Selection
- Neural network for Classification
- PCA over Attributes for Visualization and Feature Selection

Decision Tree

- Classification decision tree *fitctree* using the entire data set with default predictor values
- Mean accuracy: 0.9751

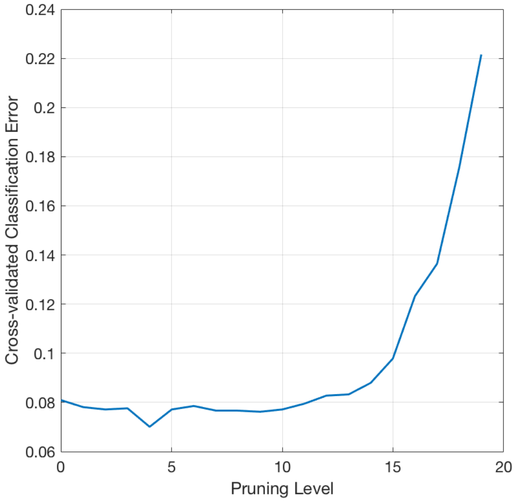
```
Default tree
-- 62 internal nodes
-- Resubtitution loss:
   0.025
-- Average loss: 0.076
```



Decision Tree: Optimization

- 10-fold cross-validation error for each subtree at various pruning levels
- Observe the best pruning level over all subtrees

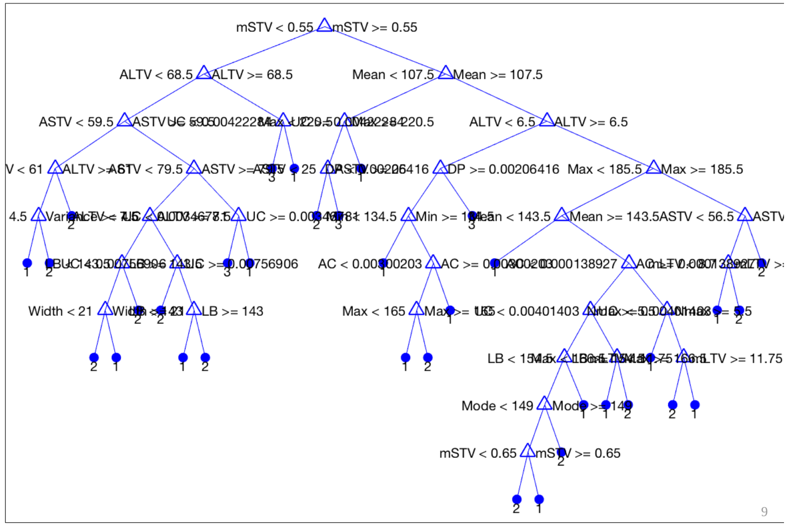
```
Optimal pruning level: 4
```



Decision Tree: Optimized Tree

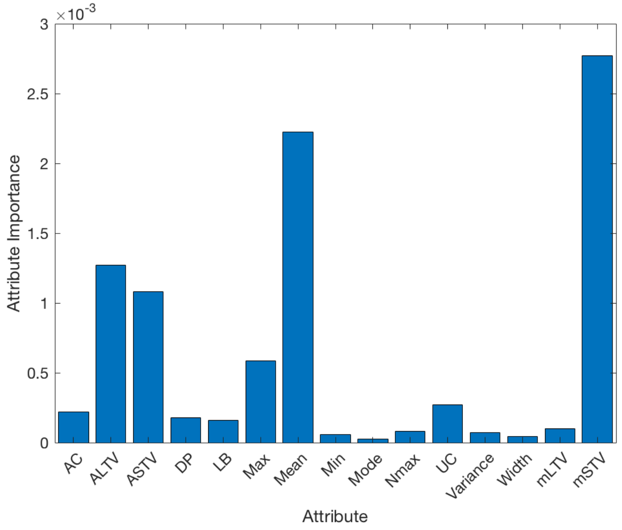
- Classification error by cross validation *cvloss*
  - Stratified partitioning to identify the optimal pruning level
- Mean accuracy: 0.9633

```
Optimal tree
-- 33 internal nodes
-- Resubtitution loss:
   0.037
-- Average loss: 0.072
```



Decision Tree: Attributes

- Attributes used for classification in optimized decision tree
- Predictors with greatest weight
  - mSTV
  - Mean
  - ALTV
  - ASTV
  - Max

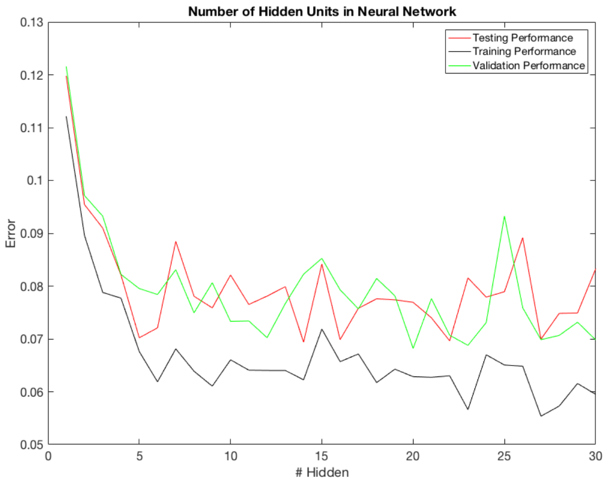


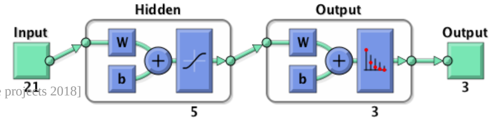
Neural Network: Data Preparation

- Processed class labels into binary representation (3 columns for 3 classes: N, S, P)
- Used a commonly-used split of training-testing data to find the best number of hidden units.
  - 70% Train, 15% Validation, 15% Test

Neural Network: Hidden Units

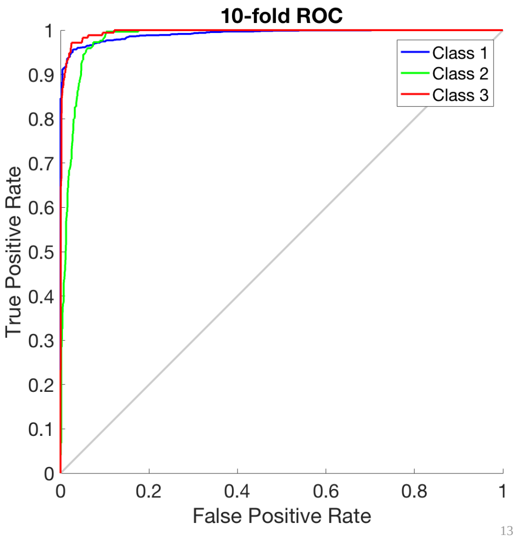
- Number of hidden units determined visually by selecting minimum of testing performance error with fewest units.
- Increased units shows increased variability in the last epoch testing and validation performance.
- First local minimum: 5 units





Neural Network: Cross-Validation

- Using the best number of hidden units, evaluated performance with k-fold cross-validation.
- 5-fold vs 10-fold showed:
  - 5 fold 3-class average AUC: 0.9895
  - 10 fold 3-class average AUC: 0.9897
  - Similar AUCs show 5 fold is sufficient for our neural network.
- Total Correct Classification (10 fold): 95.17% (4.83% incorrect)



Comparison of Classifiers

- Tried to extend this Neural Network to the 10-group classification also used in this data set.
  - These 10 classifications are a more detailed refinement of the 3-group one
- 10-Group Total Correct Classification: 84.36% (15.64% incorrect)
- This shows that the classification is worse with more groups. The classification is more complex, not as easy to predict.
  - Limitation: this comparison was done assuming the best number of hidden units for the 3-group classification.
  - Further work could be done to optimize this performance.

Compare to Literature: Decision Tree

- Huang and Hsu (2012) used discriminant analysis (DA), decision tree (DT), and artificial neural network (ANN) to evaluate fetal distress
  - DT accuracy: 86.36%
- Karabulut and Ibrikci (2014)
  - DT accuracy: 92.43%

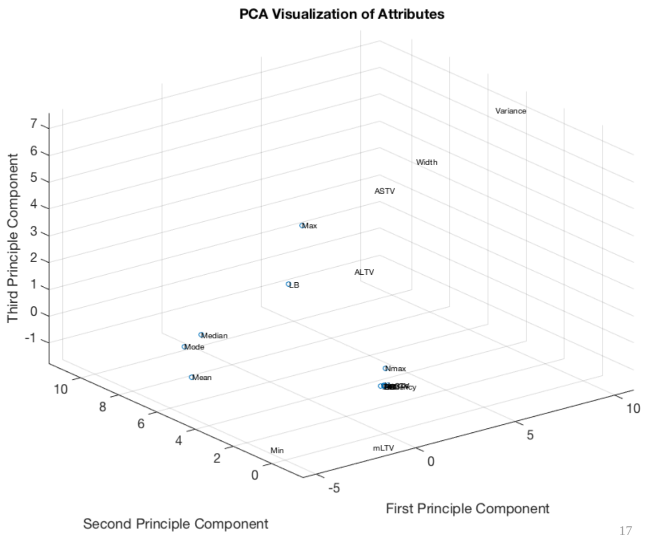
	Huang and Hsu			Our Decision Tree		
	Normal	Suspect	Pathologic	Normal	Suspect	Pathologic
Accuracy	0.932	0.681	0.528	0.966	0.969	0.992

Compare to Literature: Neural Network

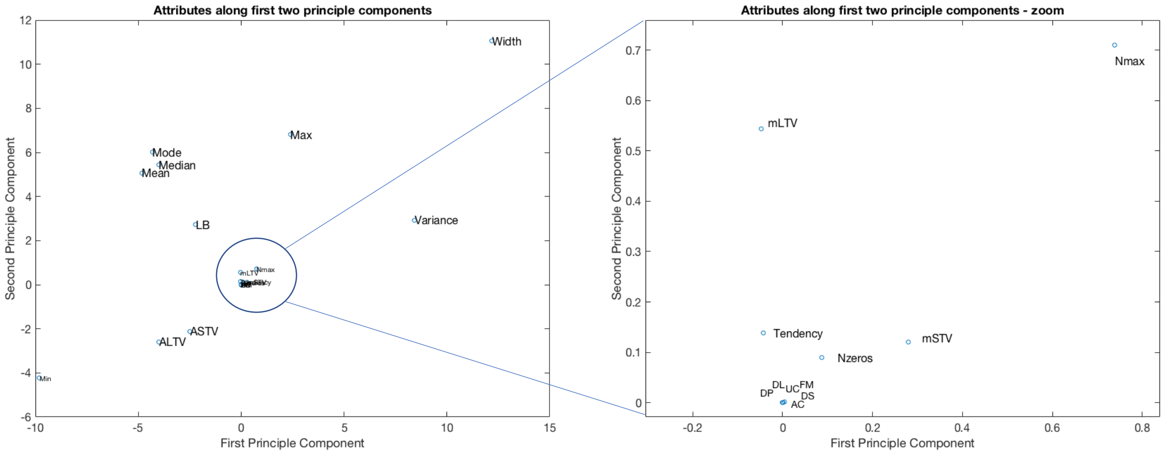
- Our accuracy: 95.17% with 5 hidden units
- Huang and Hsu (2012)
  - ANN accuracy: 97.78%
  - Used 6 hidden units
- Karabulut and Ibrikci (2014)
  - ANN accuracy: 92.09%
- Overall, we found comparable ANN accuracies with similar structures to those found in literature.

Visualizations: Attributes PCA

- PCA of the attributes revealed that the accelerations and decelerations were most closely related, followed by the uterine contractions and fetal movement.
- Wide differences were seen between attributes like the variance, variabilities, and baseline value.

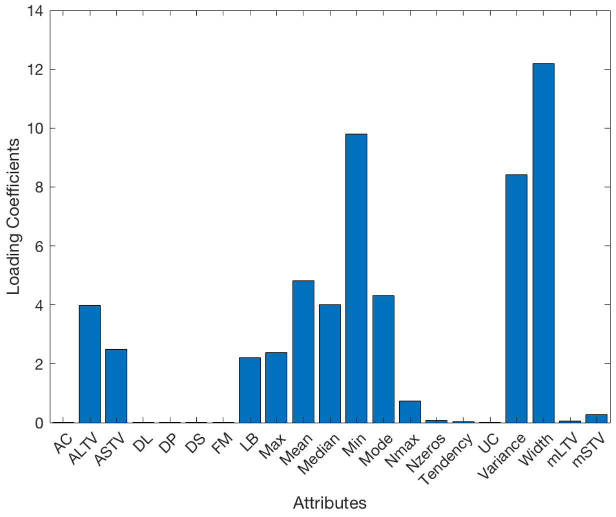


Visualization along 2 Principle Components



PCA Feature Selection

- Attributes and absolute value of loading coefficients in the first principle component
- Absolute value of loading coefficients (positive and negative) indicate strong effect on principle component



PCA Feature Selection

- Top Attributes from PCA were width, min, variance, **mean**, mode.
  - Mean, median, and mode likely to be reduced as one component of descriptive statistics.
- Top Attributes from Decision Tree were mSTV, **mean**, ALTV, ASTV, and max.
- Shared one attribute, use for future work in data dimension reduction.
  - Minimize along low variance features like accelerations.
  - Maximize along high variance features listed above.

Conclusion

- Our classification methods both performed sufficiently well, within the ranges found in other literature on the same dataset
- Neural Network slightly outperformed the accuracy of the Decision Tree but the computational time to arrive at the final tree was faster than the neural network.
- Future work should include performance of the ANN with the data after feature selection.

References

1. Alfirevic Z, Devane D, Gyte GML. (2006) Continuous cardiotocography (CTG) as a form of electronic fetal monitoring (EFM) for fetal assessment during labour. *Cochrane Database of Systematic Reviews*, Issue 3. Art. No.: CD006066. DOI: 10.1002/14651858.CD006066.
2. Narayanan, Indira & Vivio, Donna. (2016) Basic Neonatal Resuscitation: Global Landscape Analysis. 10.13140/RG.2.2.23250.15044.
3. Magenes G, Signorini MG, Arduini D. (2000) Classification of Cardiotocographic Records by Neural Networks. *Proceedings of the IEEE-INNS-ENNS International Joint Conference*, Volume 3. DOI: 10.1109/IJCNN.2000.861394.
4. Dua, D. and Karra Taniskidou, E. (2017) Cardiotocography Data Set. UCI Machine Learning Repository. Irvine, CA: University of California, School of Information and Computer Science. https://archive.ics.uci.edu/ml/datasets/cardiotocography
5. Karabulut, E.M. and Ibrikci, T. (2014) Analysis of Cardiotocogram Data for Fetal Distress Determina- tion by Decision Tree Based Adaptive Boosting Approach. Journal of Computer and Communications, 2, 32-37. http://dx.doi.org/10.4236/jcc.2014.29005