



Breast Cancer

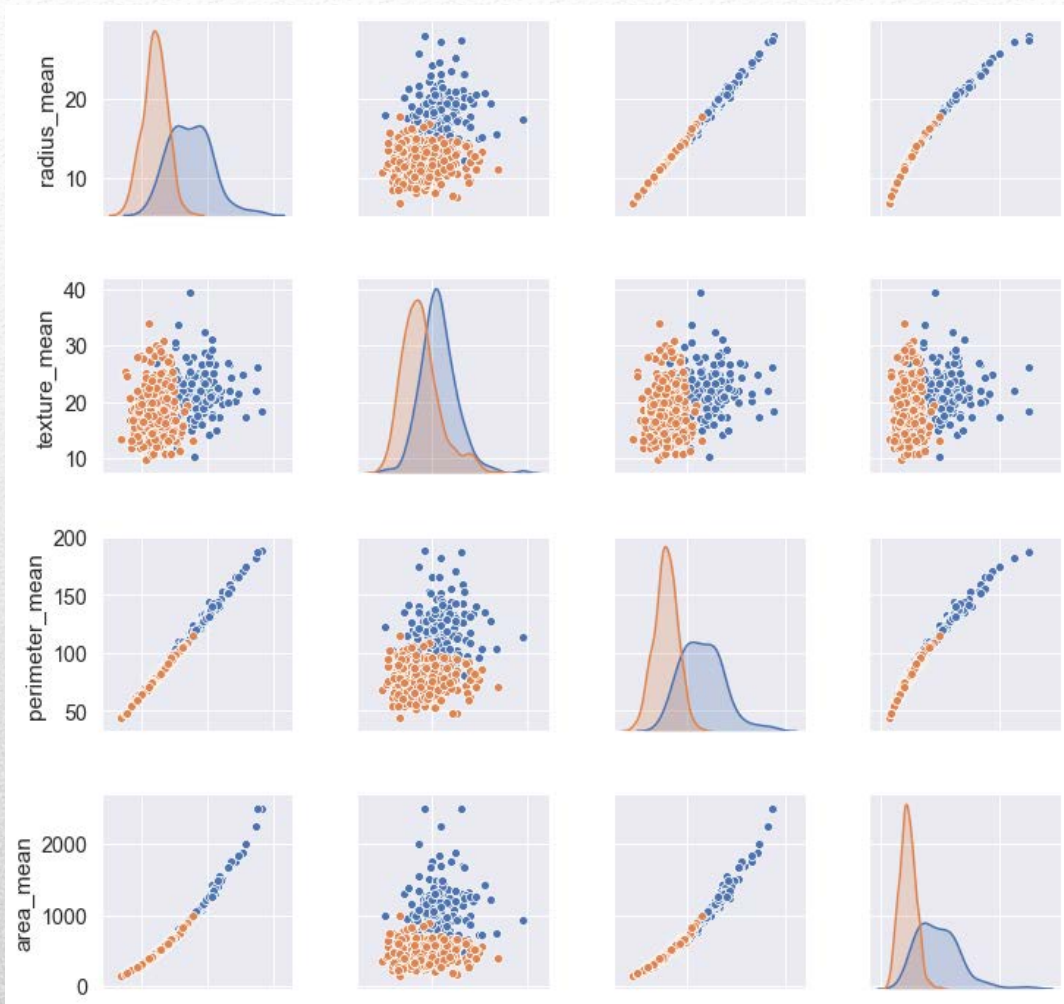
Diagnose Prediction

- The breast cancer
 - Malignant: aggressive and cancerous
 - Benign: neither aggressive nor cancerous.
- A found that a clinical diagnosis has accuracy of 90.8%, of which **3** out of **55** patient with malignant tumor were wrongly diagnosed to be benign[1]

Background

- Input: 30 features describing geometric characteristics of the tumor
- Output: diagnosis (malignant/benign)
- A supervised machine learning

Problem Statement



Visually it looks it
is easy to separate

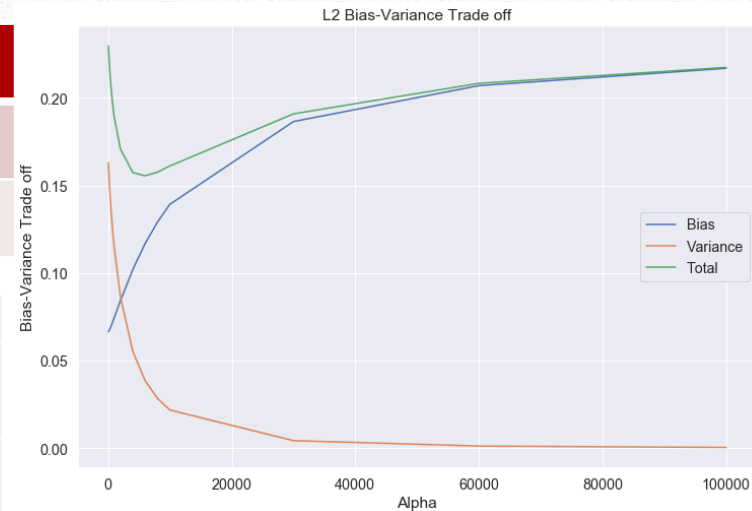
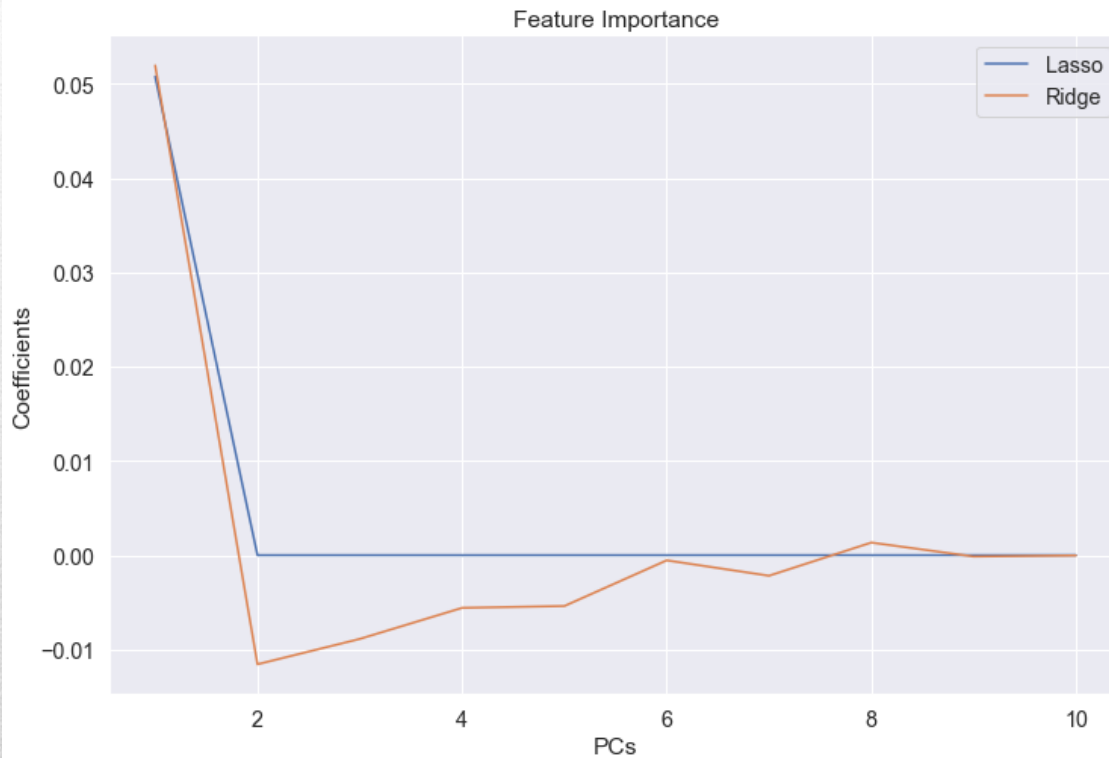
Exploratory data analysis

- Outputs n Principle Components features (much smaller than original)
 - Outlier detection (using 2 components)
 - Correlation: uncorrelated

n_components	Explained Variance Ratio
7	91.01
8	92.60
9	93.99
10	95.16

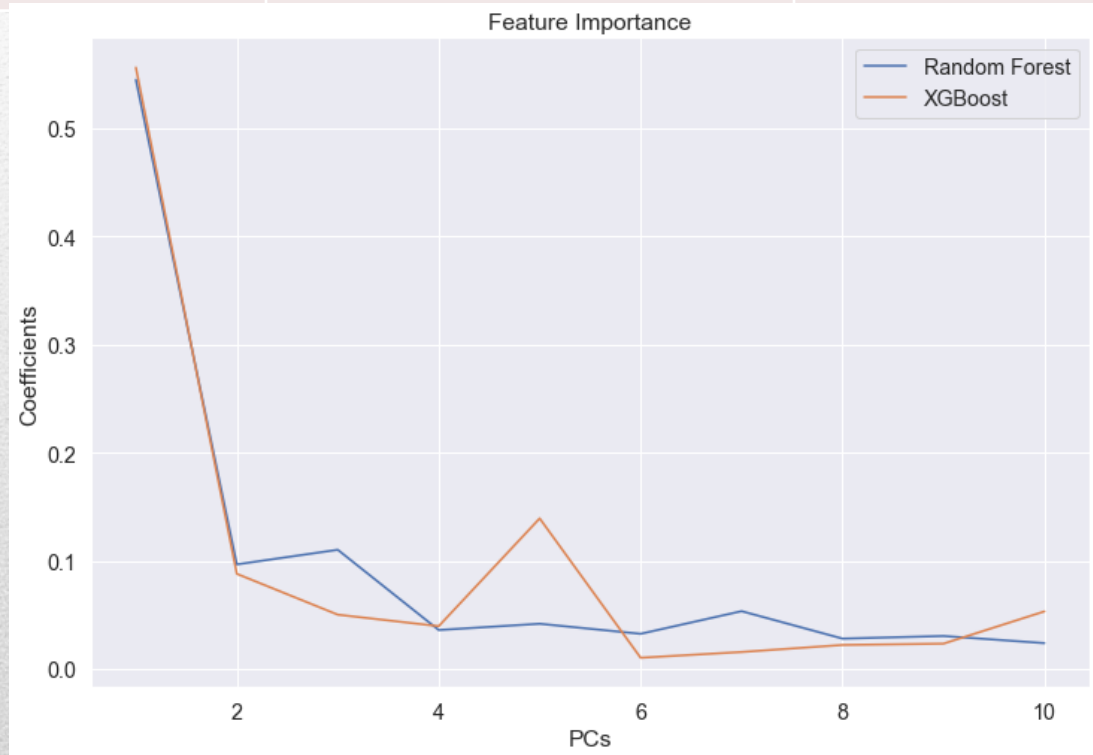
Unsupervised-PCA

	L1 Lasso	L2 Ridge
Best Alpha	0.7	6000
R2 Score	45.72%	50.04%



Supervised: L1 & L2

10 Fold Validation	Random Forest	XGBoost
Mean Recall Score	92.94%	95.88%
Recall Score Std.	5.13%	4.59%



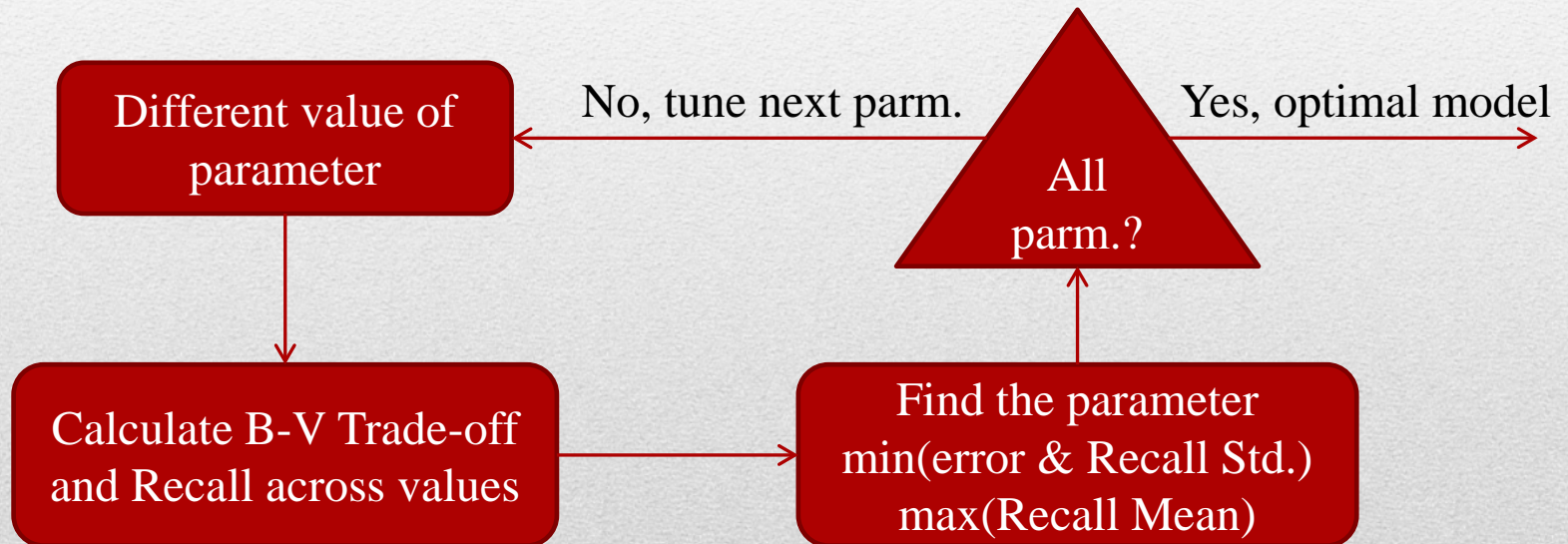
Supervised: Ensembles

Random Forest

max depth & n estimators

XGBoost

max depth, n estimators & learning rate

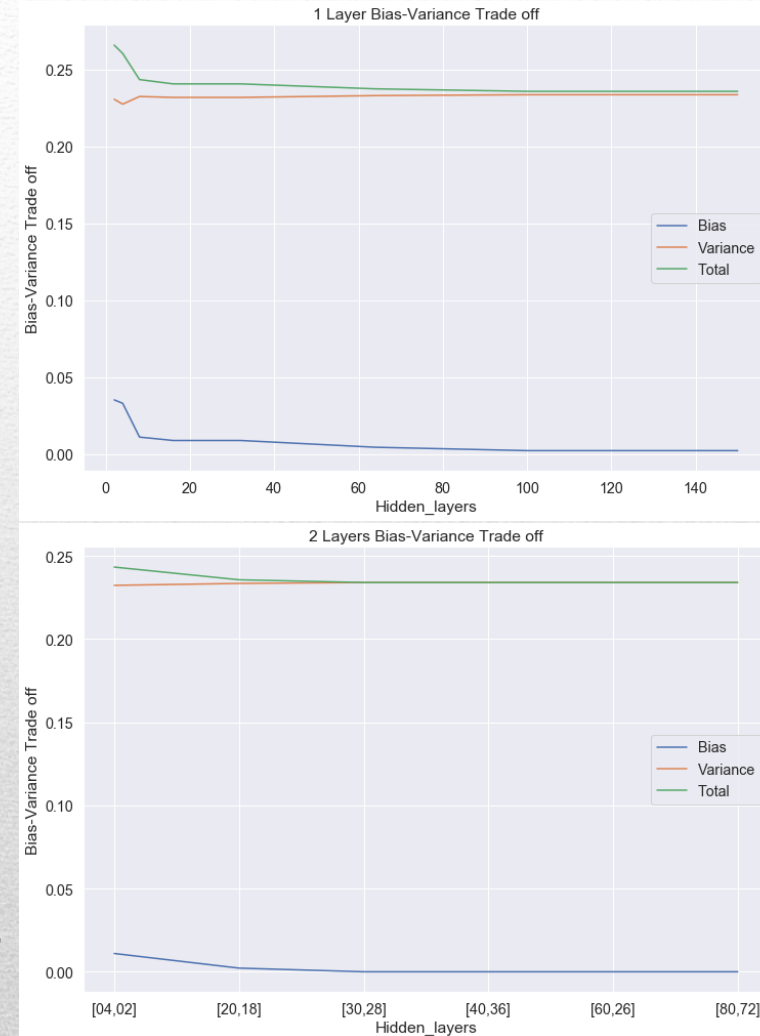


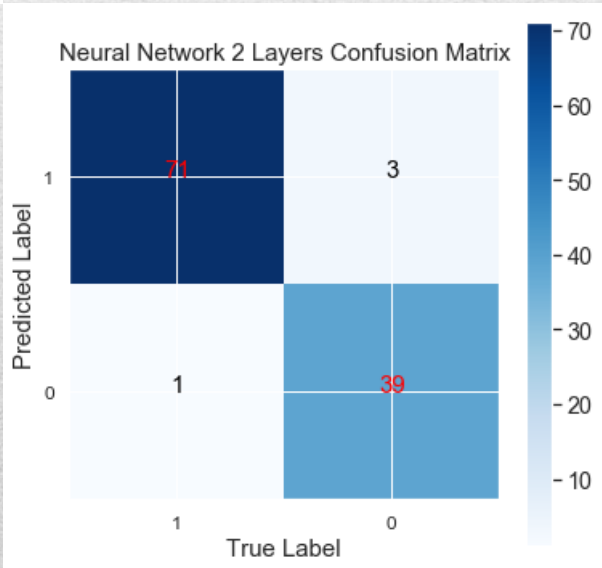
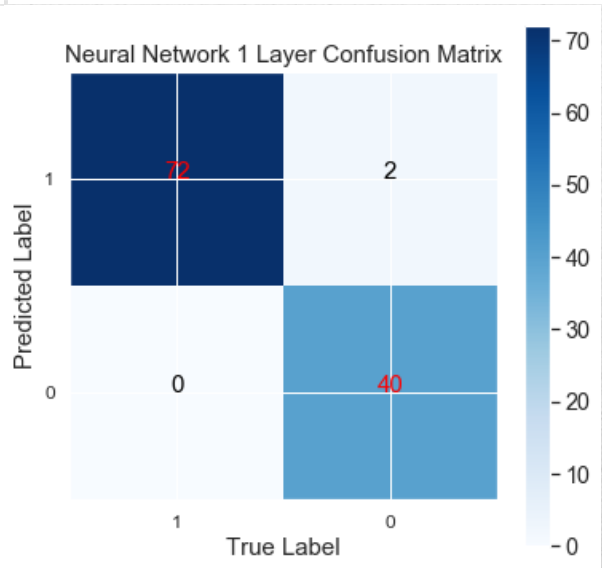
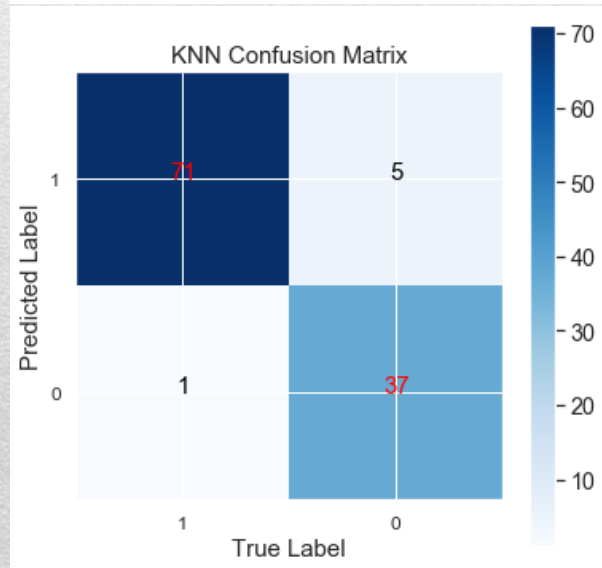
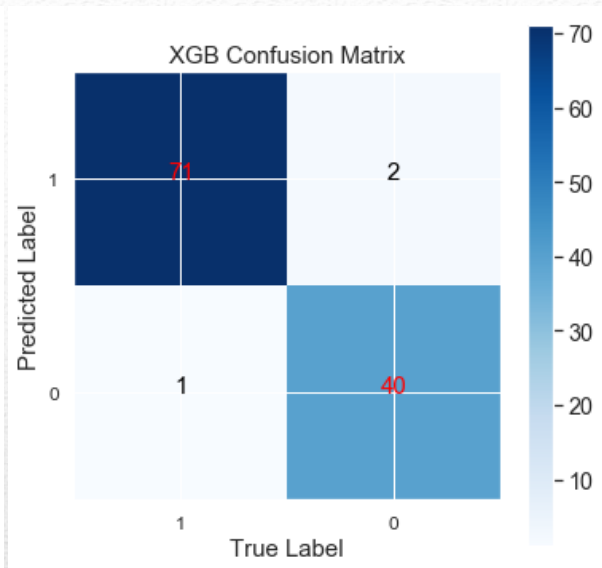
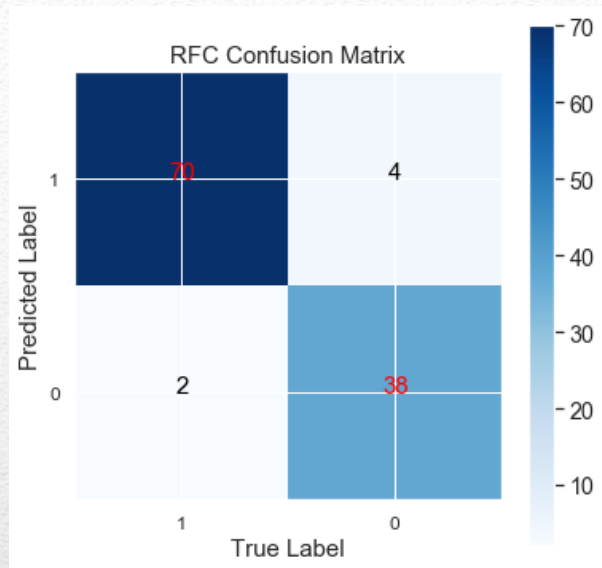
Manual Fine Tuning

- Many parameters that can be tuned
 - Max iterations
 - Hidden layers and nodes one each layer
 - Alpha
- Two layers similar power as one layer

	1 Layers	2 Layers
Mean R2	95.29%	95.88%
R2 Std.	4.40%	4.59%

Additional Models Neural Network





- 94.54%* to 95.88% to 72 out of 72
- Balance between bias and variance trade-off for constant performance
- The model can serve as a parallel method along clinics regular diagnosis

Conclusion

*3 out of 55 Malignant diagnosed as Benign: $52/55=94.54\%$



Thank you!

Questions?
