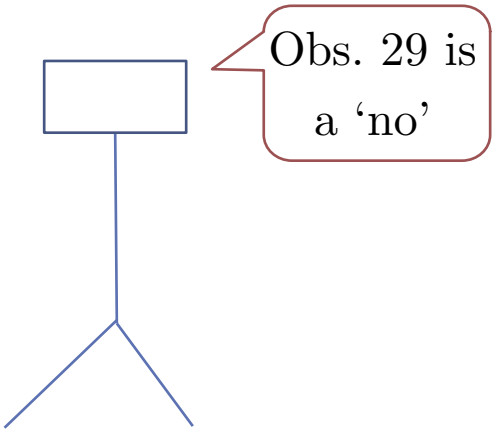
Ensemble Models

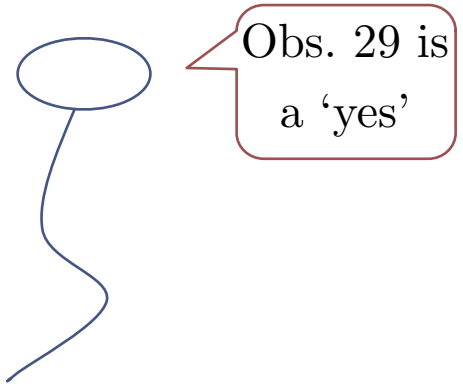
Combining models to improve performance

Simple Voting Ensemble

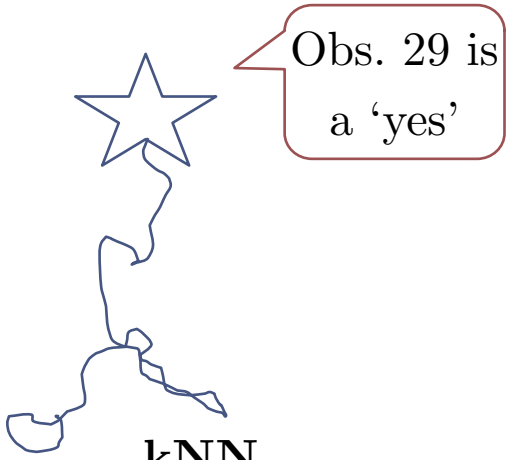
- Create a decision tree, a logistic model, and a kNN model.
- All have misclassification rate ~0.20-0.25
- If they each misclassify *different* observations...
- They may be more accurate in ensemble.



Tree Model

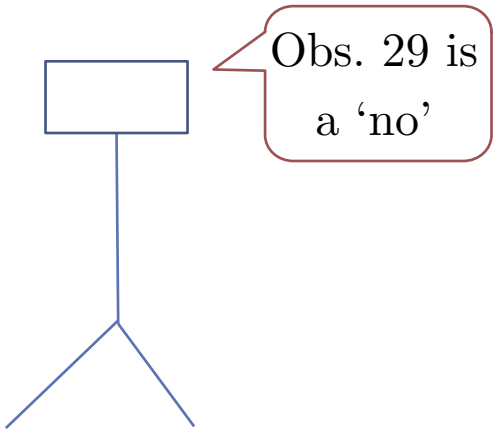


Logistic Model

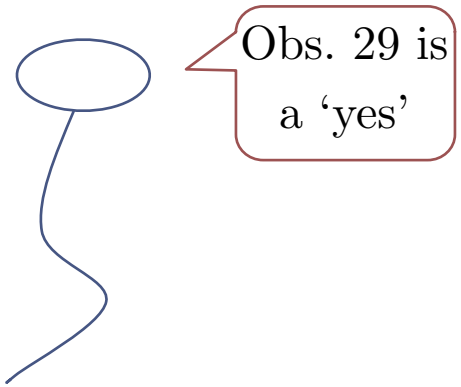


kNN Model

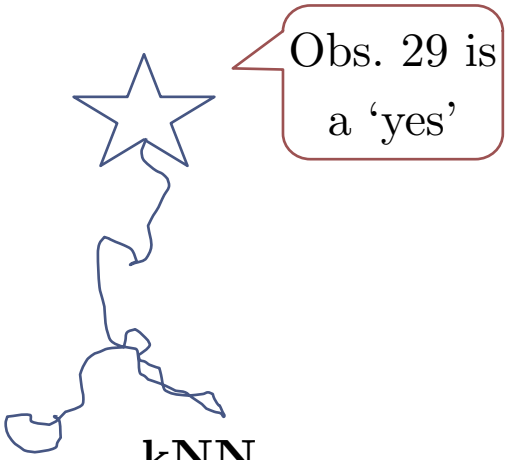
Proportion Voting Ensemble



Tree Model



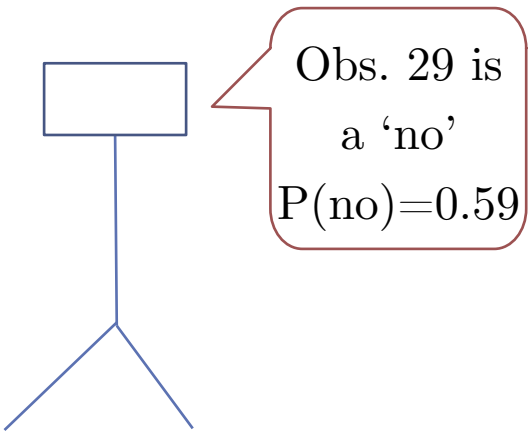
Logistic Model



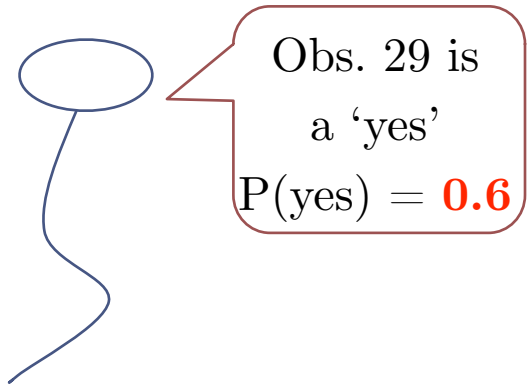
kNN Model

- An **Ensemble model** will use the input from all 3!
- Since 2 of 3 models in this ensemble say 'Yes', we'll **predict observation 29 as a 'yes' with probability** $\frac{2}{3}$.
- (SAS option Voting - Proportion)

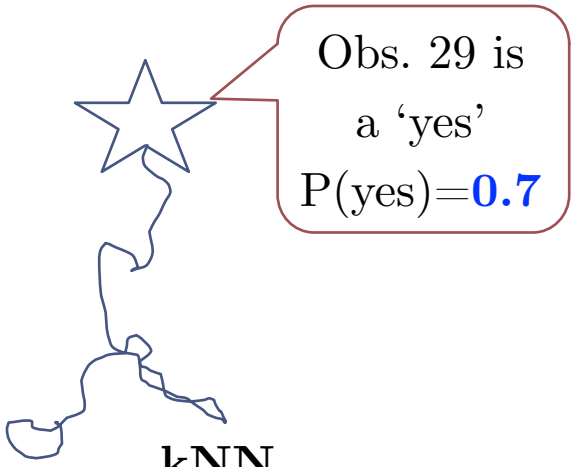
Average Voting Ensembles



Tree Model



Logistic Model



kNN Model

Predict observation 29 as a 'yes' with probability equal to average of all 'yes' probabilities *from models that say yes*:

$$\mathbf{P(yes) = (1/2)\ (0.6{+}0.7) = 0.65}$$

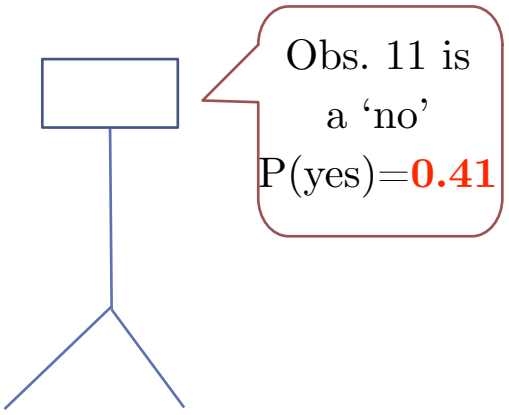
(SAS option Voting - Average)

Average Ensembles

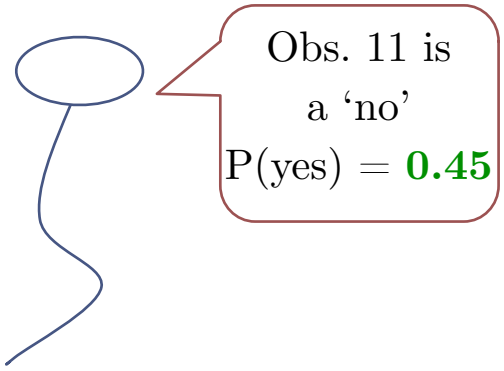
Can also consider averaging *every* model's predicted probability for the event:

- P('no') = (1/3)(0.59+0.55+0.35) = 0.496
- **P('yes') = (1/3)(0.41+0.45+0.65) = 0.503**

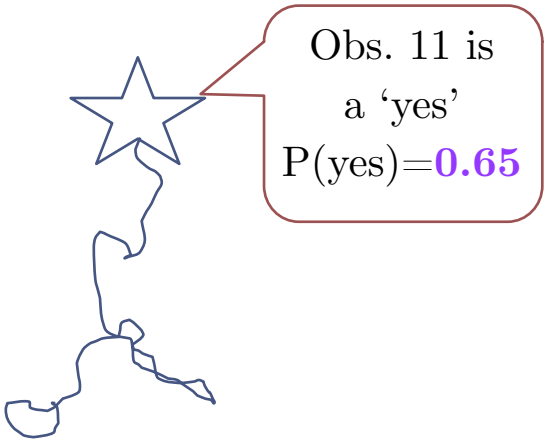
(SAS Option Average)



Tree Model



Logistic Model

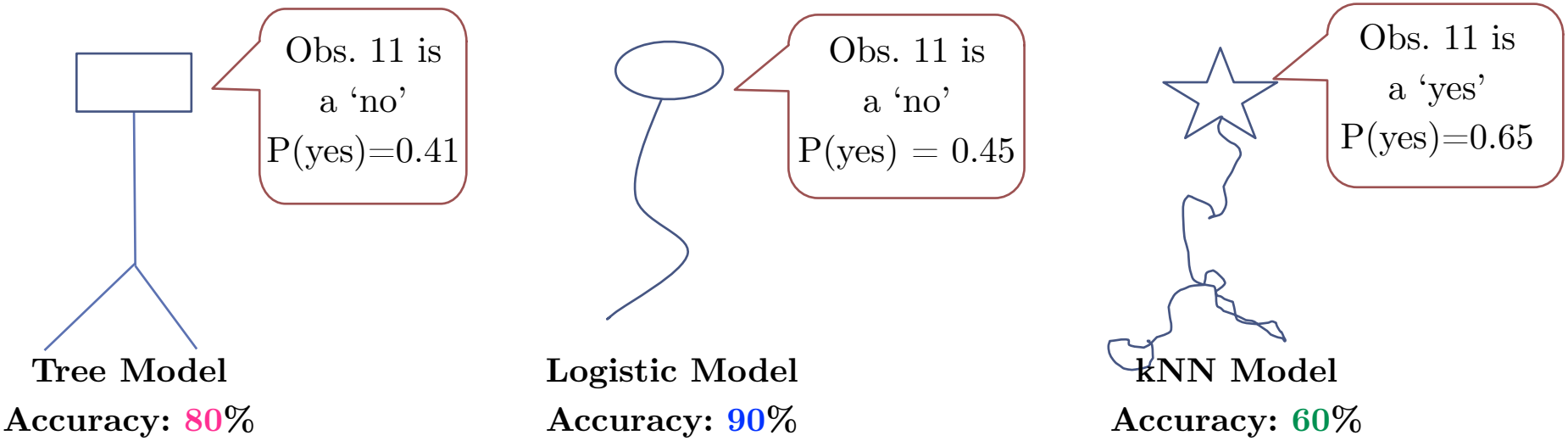


kNN Model

Weighted Vote Ensembles

May even choose to weight the predictions by the models' accuracy:

- P('no') = [**0.8***(0.59)+**0.9***(0.55)+**0.6***(0.35)] / [0.8+0.9+0.6] = 0.51
- **P('yes') = [0.8*(0.41)+0.9*(0.45)+0.6*(0.65)] / [0.8+0.9+0.6] = 0.49**

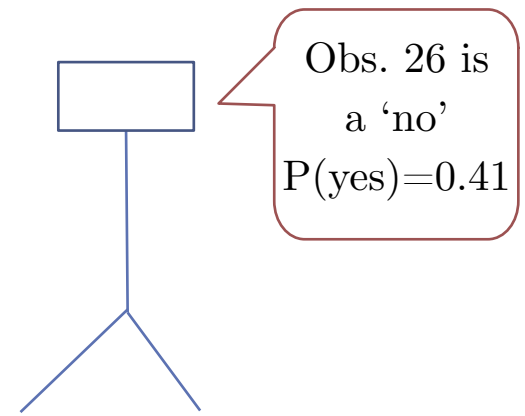


Tree Model
Accuracy: 80%

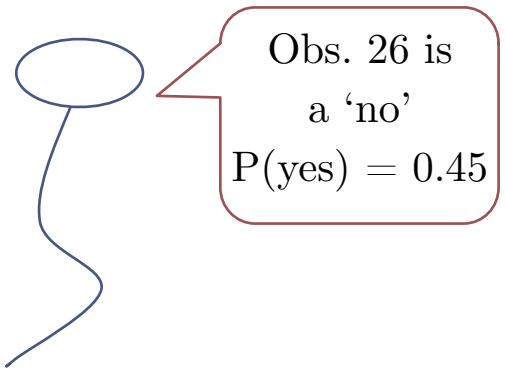
Logistic Model
Accuracy: 90%

kNN Model
Accuracy: 60%

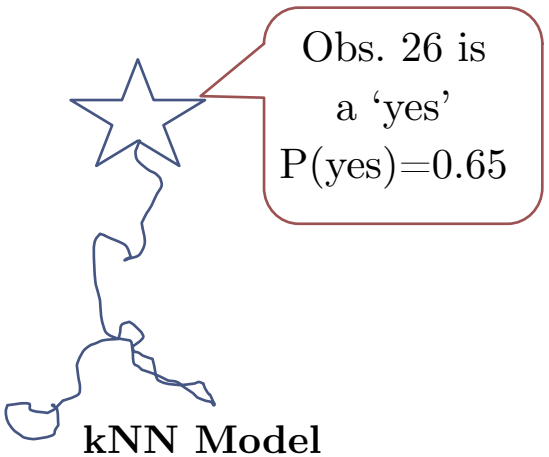
Model Stacking



Tree Model



Logistic Model



kNN Model

Observation	Model 1 Prediction	Model 2 Prediction	Model 3 Prediction	Actual Outcome
26	0.41	0.45	0.65	Yes

Combine model predictions with original inputs into NEW model (of any/all varieties) as input variables in an attempt to let a model decide if there are certain circumstances allowing each model in the ensemble works best.

Introduction to Model Builder Studio

• • •

(SAS's Enterprise Miner Reborn in the Cloud)

Ensembling Models in Viya's Model Builder