* Describe one real-world application in industry where the model can be applied.
* What are the strengths of the model; when does it perform well?
* What are the weaknesses of the model; when does it perform poorly?
* What makes this model a good candidate for the problem, given what you know about the data?

Gaussian Naive Bayes (GaussianNB)

* GaussianNB can be applied to document classification and spam email filtering.
* The strengths of GaussianNB are that it requires small size of training data to determine necessary parameters and is very fast, not suffering “curse of dimensionality” (SLUG).
* Although GaussianNB is a good classifier, it can not provide good estimates of probabilities (SLUG), which is one of the weaknesses. Another weakness is that GaussianNB requires a strong conditional independence assumption on the attributes in the model (Class notes; SLUG).
* As our problem has many input variables and observations, GaussianNB with high efficiency and capability of handling high dimensionality is a good candidate.

Support Vector Machines (SVM)

Ensemble Methods (AdaBoost)

Reference:

Class notes.

Scikit-Learn User Guide (SLUG), Release 0.20.dev0.