* 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. For example, Gaussian NB takes term-document matrix of emails as the input and classify emails into spam and non-spam.
* 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 classification 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)

* SVM can be used in the imaging application of detecting human face. SVM discovers a squared boundary around face and classifies the images as with-face or without-face.
* SVM is effective in high dimensional spaces and memory efficient; it can adopt different Kernel functions for the dicision function (SLUG).
* The weakness of SVM is that its accuracy in terms of over-fitting is sensitive to the Kernel functions and regularization term if there are too many features (SLUG). Another weakness of SVM is that it do not provide probability estimates (SLUG).
* Similar to GaussianNB, SVM with high efficiency and capability of handling high dimensionality is a good candidate, as our classification problem has many input variables and observations.

Ensemble Methods (AdaBoost)

* Industrial applications of AdaBoost includes document classification and face detection.
* As the strength, AdaBoost is efficient and setting parameters for AdaBoost is easy (Class notes, ESL). As a kind of AdaBoost, boosting tree methods is highly accurate and is capable of handling irrelevant features (Class notes, ESL).
* As the weakness, AdaBoost is sensitive to data’s noise and needs enough data for fitting (Class notes, ESL).
* Our classification problem has enough data and many features, part of which may be irrelevant for fitting. Hence, AdaBoost (Boosting tree) is a good candidate for our problem.

Gradient Tree Boosting

* Gradient Tree Boosting can be considered as the off-the-shelf procedure that is the first choice of many industrial applications, such as document classification and house price prediction (SLUG).
* The strength of Gradient Tree Boosting includes natural handling of complex data features (i.e., mixed types), great predictive power, robustness to outliers in output space, and capability of interpretation (SLUG; ESL).
* The weakness of Gradient Tree Boosting is that it lacks the power of scalability, i.e., parallel computation(SLUG).
* Our classification problem has enough data and many features, which have mixed data types. Hence, Gradient Tree Boosting is a good candidate for our problem.

Reference:

Class notes.

Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. The elements of statistical learning. Vol. 1. New York: Springer series in statistics, 2001. (ESL)

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