The Art of Machine Learning for Predictive Data Analytics

John D. Kelleher and Brian Mac Namee and Aoife D'Arcy

1. **Different Perspectives on Prediction Models**
2. **Choosing a Machine Learning Approach**
   - Matching Machine Learning Approaches to Projects
   - Matching Machine Learning Approaches to Data
3. **Your Next Steps**

- Predictive data analytics projects use machine learning to build models that capture the relationships in large datasets between descriptive features and a target feature.
- Machine Learning $\approx$ inductive learning
    1. a model learned by induction is not guaranteed to be correct.
    2. learning cannot occur unless the learning process is biased in some way.

- *En masse* all of the questions that must be answered to successfully complete a predictive data analytics project can seem overwhelming.
- This is why we recommend using the **CRISP-DM** process to manage a project through its lifecycle.

CRISP-DM	Open Questions	Chapter
Business Understanding	*What is the organizational problem being addressed? In what ways could a prediction model address the organizational problem? Do we have situational fluency? What is the capacity of the organization to utilize the output of a prediction model? What data is available?*	Chapter 2
Data Understanding	*What is the prediction subject? What are the domain concepts? What is the target feature? What descriptive features will be used?*	Chapter 2

CRISP-DM	Open Questions	Chapter
Data Preparation	*Are there data quality issues? How will we handle missing values? How will we normalize our features? What features will we include?*	Chapter 3
Modelling	*What types of models will we use? How will we set the parameters of the machine learning algorithms? Have underfitting or overfitting occurred?*	Chapters 4, 5, 6 and 7

CRISP-DM	Open Questions	Chapter
Evaluation	*What evaluation process will we follow? What performance measures will we use? Is the model fit for purpose?*	Chapter 8
Deployment	*How will we continue to evaluate the model after deployment? How will the model be integrated into the organization?*	Section 8.4.6 and Chapters 9 and 10

Different Perspectives on Prediction Models

$$H(t, \mathcal{D}) = -\sum_{l \in levels(t)} (P(t = l) \times log_2(P(t = l))) \tag{1}$$

$$dist(\mathbf{q}, \mathbf{d}) = \sqrt{\sum_{i=1}^{m} (\mathbf{q}[i] - \mathbf{d}[i])^2} \tag{2}$$

$$P(t = l|\mathbf{q}) = \frac{P(\mathbf{q}|t = l)P(t = l)}{P(\mathbf{q})} \tag{3}$$

$$L_2(\mathbb{M}_{\mathbf{w}}, \mathcal{D}) = \frac{1}{2} \sum_{i=1}^{n} (t_i - \mathbb{M}_{\mathbf{w}}(\mathbf{d}_i))^2 \tag{4}$$

parametric versus non-parametric

- generally describes whether the the size of the **domain representation** used to define a model is solely determined by the number of features in the domain or if it is affected by the number of instances in the dataset.
- parametric model the size of the domain representation is independent of the number of instances in the dataset (e.g., naive Bayes', regression models)
- non-parametric model the number of parameters used by the model increases as the number of instances increases (e.g., decision trees)

generative versus discriminative

- A model is generative if it can be used to generate data that will have the same characteristics as the dataset from which the model was produced.
- A discriminative models learn the boundary between classes rather than the characteristics of the distributions of the different classes.

- Generative and discriminative models attempt to learn different things!

A generative model works by:

1. learning $P(\mathbf{d}|t_l)$ (class conditional densities) and $P(t_l)$;
2. then using Bayes' theorem to compute $P(t_l|\mathbf{d})$;
3. applying a decision rule over the class posteriors to return a target level.

A discriminative model works by:

1. learning the class posterior probability $P(t_l|\mathbf{d})$ directly from the data
2. and then applying a decision rule over the class posteriors to return a target level.

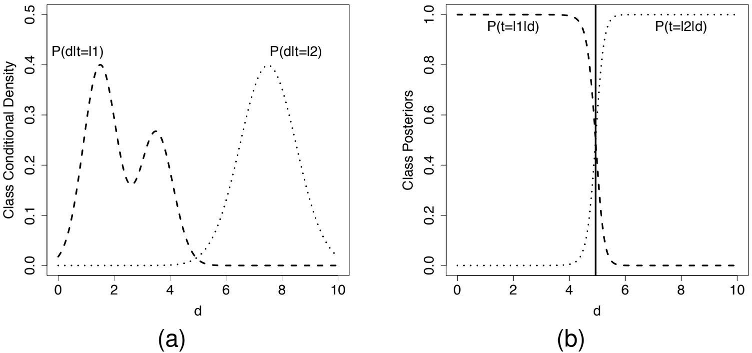


Figure: (a) The class conditional densities for two classes (*'l1'*,*'l2'*) with a single descriptive feature **d**. (b) The class posterior probabilities plotted for each class for different values of **d**. $P(t = \textit{'l1'} | \mathbf{d})$ is not affected by the multimodal structure of the corresponding class conditional density $P(\mathbf{d} | t = \textit{'l1'})$. Based on Figure 1.27 from (Bishop, 2006).

Table: A taxonomy of models based on the parametric versus non-parametric and generative versus discriminative distinctions.

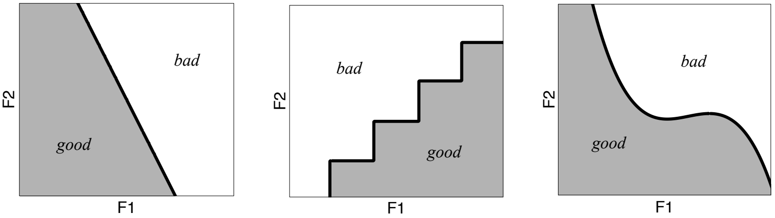
Model	Parametric/ Non-Parametric	Generative/ Discriminative
k nearest neighbor	Non-Parametric	Generative
Decision Trees	Non-Parametric	Discriminative
Bagging/Boosting	Parametric[1]	Discriminative
Naive Bayes	Parametric	Generative
Bayesian Network	Parametric	Generative
Linear Regression	Parametric	Discriminative
Logistic Regression	Parametric	Discriminative
SVM	Non-Parametric	Discriminative

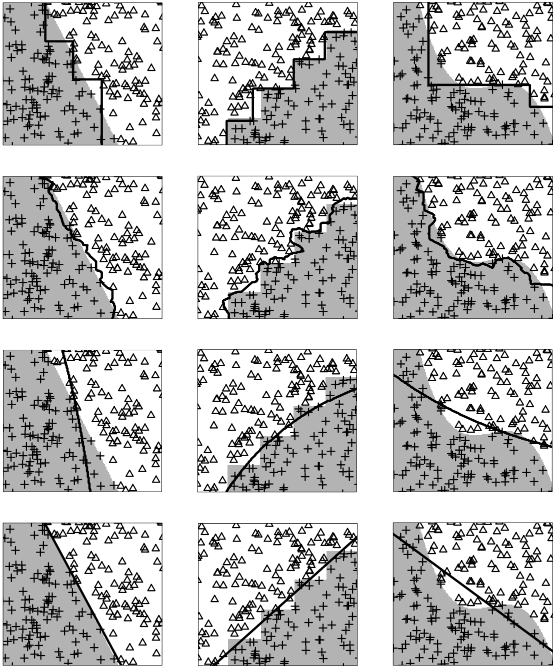
[1] While the individual models in an ensemble could be non-parametric (for example when decision trees are used) the ensemble model itself is considered parametric.

Choosing a Machine Learning Approach

- there is not one best approach that always outperforms the others; no free lunch theorem.
- We can see the assumptions encoded in each algorithm reflected in the distinctive characteristics of the decision boundaries that they learn for categorical prediction tasks.

Data Sets





- Typically choose a number of different approaches and to run experiments to evaluate which best the particular project.
- There are two questions to consider in the selection of a set of initial approaches:
  1. Does a machine learning approach match the requirements of the project?
  2. Is the approach suitable for the type of prediction we want to make and the types of descriptive features we are using?

Project Requirements

- Accuracy
- Prediction speed
- Capacity for retraining
- Interpretability

Data Considerations

- continuous target $\rightarrow$ error based models
- categorical target $\rightarrow$ information and probability models
- continuous descriptive features $\rightarrow$ (+cat. target) similarity based models / (+cont. target) error based models
- categorical descriptive features $\rightarrow$ information and probability models
- lots of descriptive features (curse of dimensionality) $\rightarrow$ feature selection

Your Next Steps

- The easy part of a predictive data analytics project is building the models.
- What makes predictive data analytics difficult, but also fascinating, is figuring out how to answer all the questions that surround the modelling phase of a project.
- Intuition, experience, and experimentation!

Key tasks for an analyst

- become situationally fluent so that we can converse with experts in the application domain;
- explore the data to understand it correctly;
- spend time cleaning the data;
- think hard about the best ways to represent features;
- spend time designing the evaluation process correctly.

- Machine learning is a huge topic.
- There are lots of topics we haven't covered: **deep learning**, **graphical models**, **multi-label classification**, **association mining**, **clustering**, ...
- But, we hope this course has given you the knowledge and skills that you will need to explore machine learning yourself.

1. **Different Perspectives on Prediction Models**

2. **Choosing a Machine Learning Approach**
   - Matching Machine Learning Approaches to Projects
   - Matching Machine Learning Approaches to Data

3. **Your Next Steps**