Types of Machine Learning

Machine Learning
CSx824/ECEx242
Bert Huang
Virginia Tech

Outline

- Types of learning settings
  - Supervised learning
  - Unsupervised learning
- Types of learning algorithms
  - Batch learning
  - Online learning

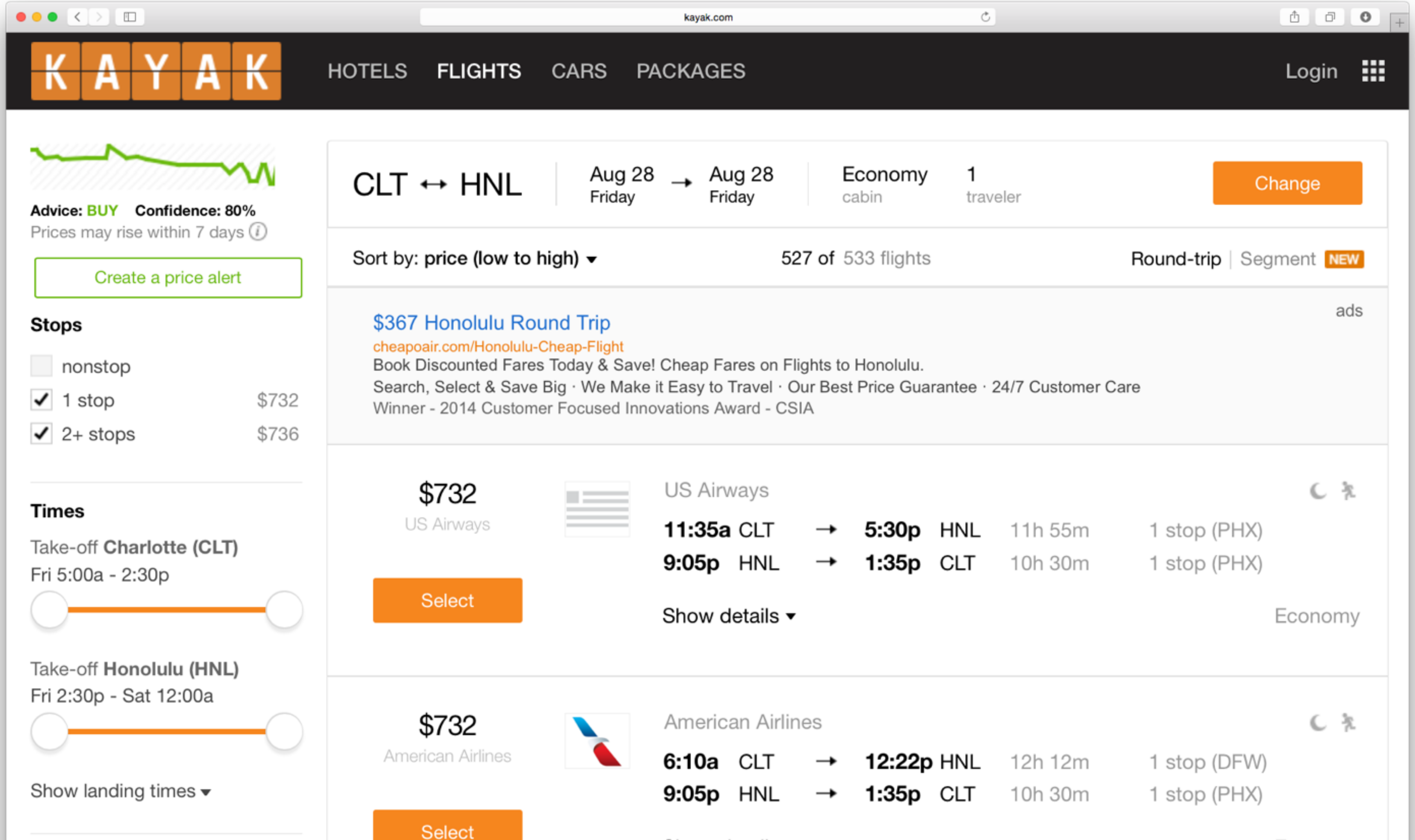
1st Learning Setting

- Draw data set $D = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$ from distribution $\mathbb{D}$
- Algorithm A learns hypothesis $h \in H$ from set H of possible hypotheses $A(D) = h$
- We measure the quality of h as the expected **loss**: $\underset{(x,y) \in \mathbb{D}}{E} [\ell(y, h(x))]$
  - This quantity is known as the **risk**
  - E.g., loss could be the Hamming loss $\ell_{\text{Hamming}}(a, b) = \begin{cases} 0 & \text{if } a = b \\ 1 & \text{otherwise} \end{cases}$

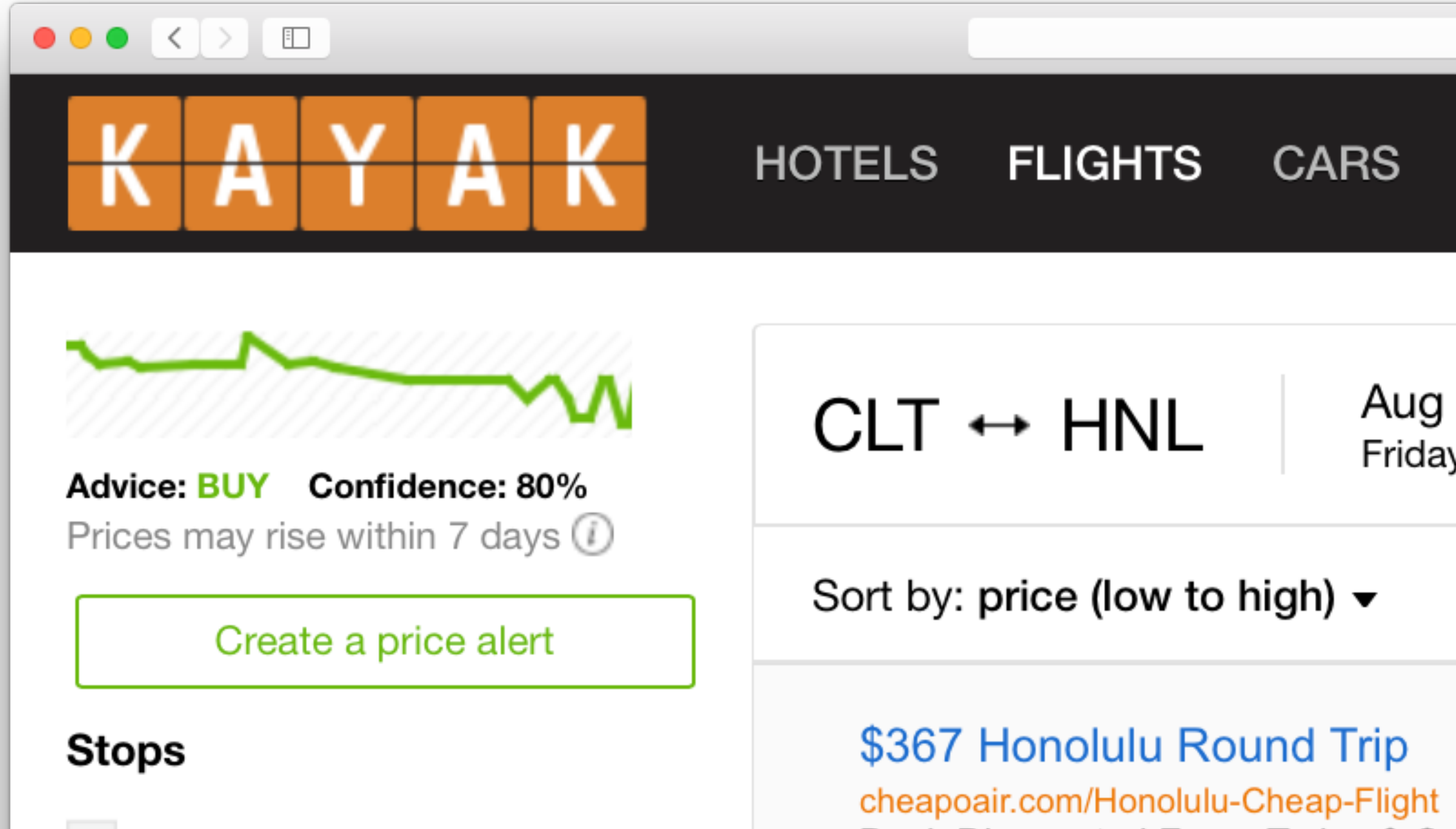
Example: Digit Classification

http://ufldl.stanford.edu/housenumbers/

Example: Airline Price Prediction



Example: Airline Price Prediction



Batch Supervised Learning

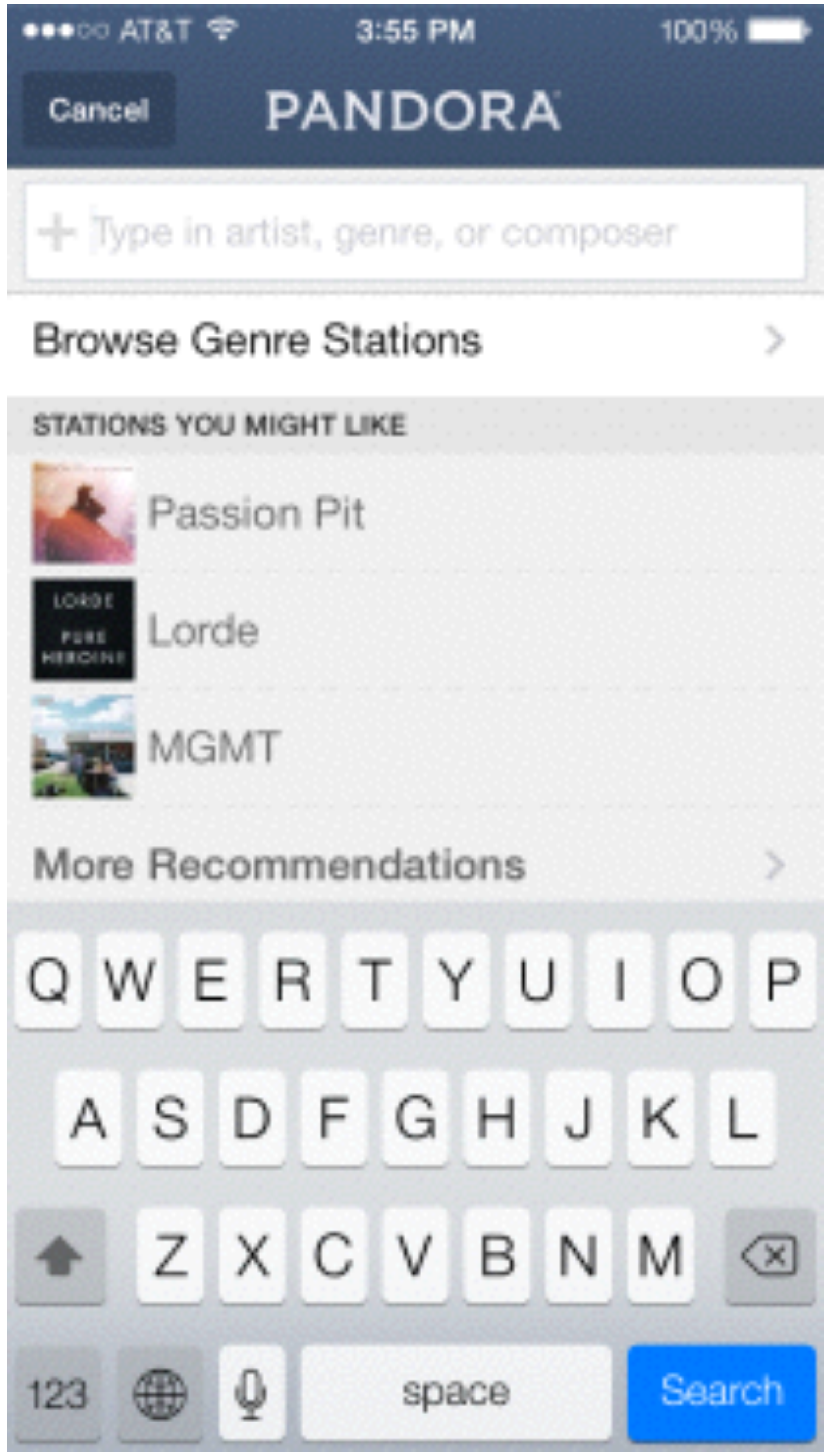
- Draw data set $D = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$ from distribution $\mathbb{D}$
- Algorithm A learns hypothesis $h \in H$ from set H of possible hypotheses $A(D) = h$
- We measure the quality of h as the expected **loss**: $\underset{(x,y)\in\mathbb{D}}{E}[\ell(y, h(x))]$
  - This quantity is known as the **risk**
  - E.g., loss could be the Hamming loss $\ell_{\mathrm{Hamming}}(a, b) = \begin{cases} 0 & \text{if } a = b \\ 1 & \text{otherwise} \end{cases}$

classification

Online Supervised Learning

- In step ***t***, draw data point ***x*** from distribution $\mathbb{D}$
- Current hypothesis ***h*** guesses the label of ***x***
- Get true label from oracle ***O***
- Pay penalty if ***h*(*x*)** is wrong (or earn reward if correct)
- Learning algorithm updates to new hypothesis based on this experience
    - Does not store history

Example: Recommendation



Learning Settings

- Supervised or unsupervised (or semi-supervised, weakly supervised, transductive…)
- Online or batch (or reinforcement…)
- Classification, regression
  - (or structured output, clustering, dimensionality reduction…)
- Parametric or non-parameteric

Functional Perspective

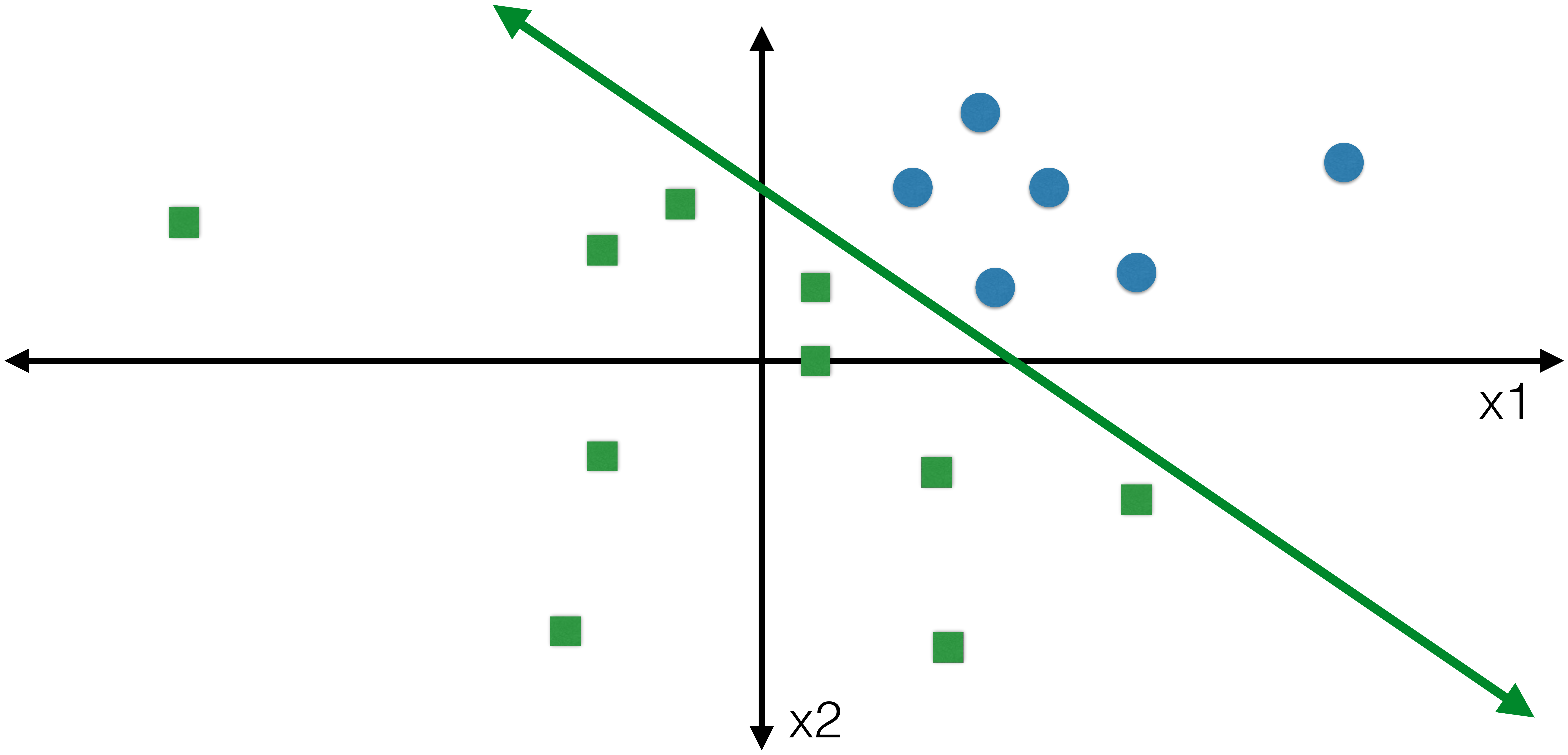
Input	Learning Setting
Batch of Data Points with Labels	Batch Supervised Learning
Batch of Data Points	Batch Unsupervised Learning
Data Point(s) and Previous Model	Online Supervised Learning

Parametric vs. Non-Parametric

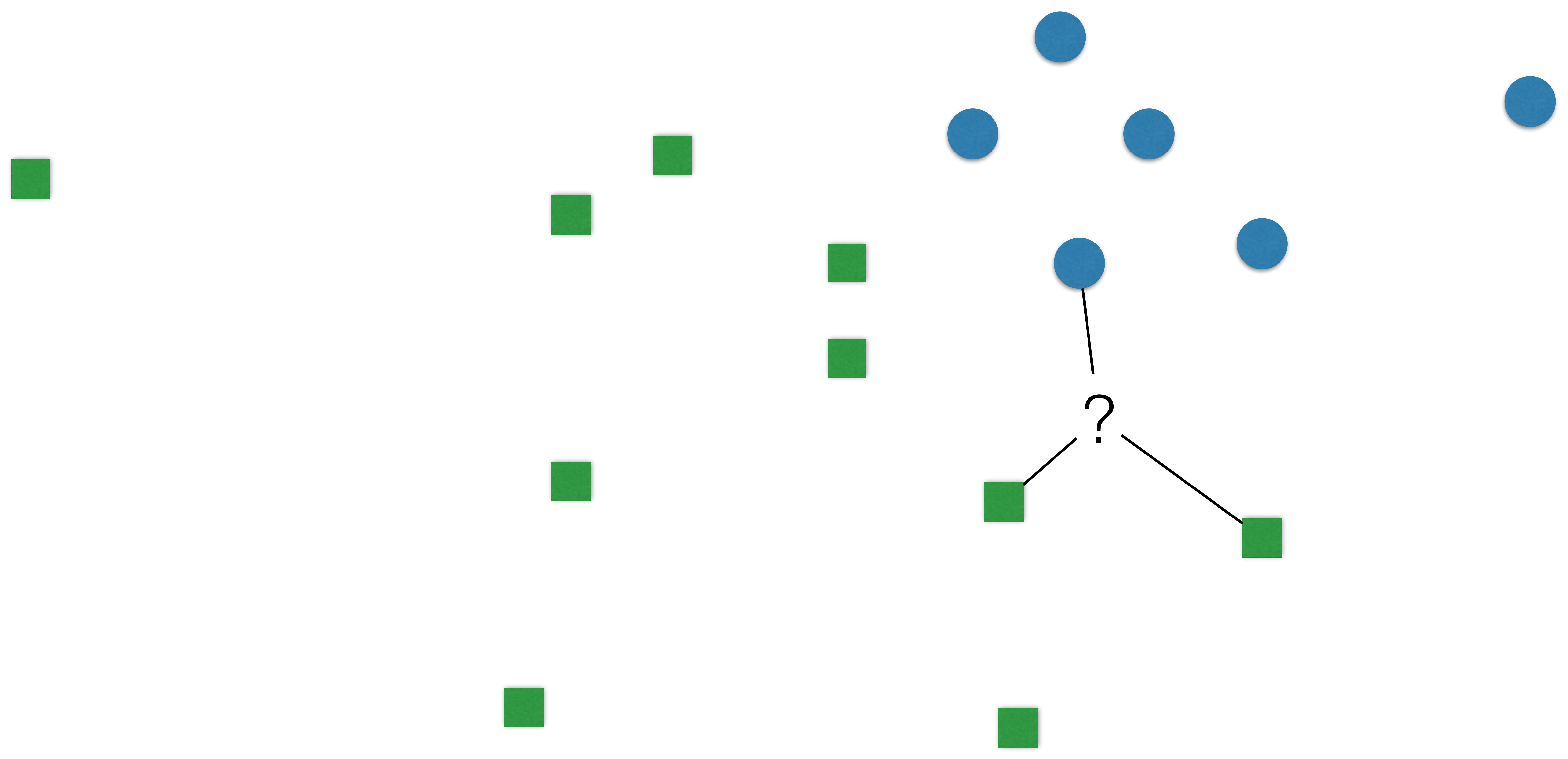
$$A(\mathcal{D}) = h \qquad\qquad h \in H$$

- If $h \in H$ identifiable by finite-dimensional vector: parametric
  - e.g., linear classifier $h_{\boldsymbol{\theta}}(\mathbf{x}) = \textsf{sign}\left(\boldsymbol{\theta}^\top \mathbf{x}\right)$
- If $h \in H$ has a flexible number of parameters: non-parametric
  - e.g., k-nearest neighbor

Parametric Classifier: Linear



Non-Parametric Classifier: K-Nearest Neighbors



Learning Settings and Probability

	discriminative	generative
Supervised	$p(y\|x)$ $p(y\|x;\theta)$	$p(y, x)$ $p(y, x; \theta)$
Unsupervised		$p(x)$ $p(x; \theta)$

θ finite dimensional: parametric

non-parametric otherwise

Concepts

- Supervised and unsupervised learning
- Parametric and non-parametric learning algorithms
- Online and batch learning
- Discriminative and generative
- Output of models: classification and regression

Discussion Questions

- How do you characterize different machine learning algorithms you know about?
- Are learning-algorithm attributes independent? Are there combinations of attributes that fit well together or don't fit well?