Introduction to Machine Learning

COMPSCI 371D — Machine Learning
Outline

1. Unsupervised Learning
2. Drawings and Intuition in Higher Dimensions
3. Classification through Regression
4. Linear Separability
Unsupervised Learning

Parenthesis: Supervised vs Unsupervised

- **Supervised**: Train with \((x, y)\)
  - Classification: Hand-written digit recognition
  - Regression: Median age of YouTube viewers for each video

- **Unsupervised**: Train with \(x\)
  - Clustering: Group customers by similar tastes to focus advertising
  - Dimensionality reduction: Which dimensions contain most of the variation?

We will *not* cover unsupervised learning

[Image from cw.fel.cvut.cz]
Drawings Help Intuition

[Image of a scatter plot with blue and red points]

COMPSCI 371D — Machine Learning
Introduction to Machine Learning

4/12
Classifiers as Partitions of $X$

- $\hat{y} = h(x)$ for $\hat{y} \in Y$, a categorical set
- $X_y \overset{\text{def}}{=} h^{-1}(y)$ \textit{partitions} $X$ (not just $T$!)
- Classifier $= \text{partition}$
- $S = h^{-1}(\text{red square}), \ C = h^{-1}(\text{blue circle})$
Intuition Often Fails in Many Dimensions

- Gray parts dominate when $d \to \infty$
- Distance from center to corners diverges when $d \to \infty$
A classifier partitions $X \subset \mathbb{R}^d$ into sets, one per label in $Y$.

How do we represent sets $\subset \mathbb{R}^d$? How do we work with them?

We’ll see a couple of ways: nearest-neighbor classifier, decision trees.

These methods have a strong geometric flavor.

Beware of our intuition!

Another technique: score-based classifiers

*i.e.*, classification through regression.

Examples: linear classifiers, support vector machines, neural networks.
Score-Based Classifiers

- $s = 0$ defines the decision boundaries
- $s > 0$ and $s < 0$ defines the (two) decision regions

[Figure adapted from Wei et al., *Structural and Multidisciplinary Optimization*, 58:831–849, 2018]
Score-Based Classifiers

- Threshold some score function $s(x)$:
- Example: 's' (red squares) and 'c' (blue circles)

Correspond to two sets $S \subseteq X$ and $C = X \setminus S$

If we can estimate something like $s(x) = \mathbb{P}[x \in S]$

$$h(x) = \begin{cases} 's' & \text{if } s(x) > 1/2 \\ 'c' & \text{otherwise} \end{cases}$$
Classification through Regression

• If you prefer 0 as a threshold, let

\[ s(x) = 2\mathbb{P}[x \in S] - 1 \in [-1, 1] \]

\[
h(x) = \begin{cases} 
's' & \text{if } s(x) > 0 \\
'c' & \text{otherwise}
\end{cases}
\]

• Scores are convenient even without probabilities, because they are easy to work with

• We implement a classifier \( h \) by building a regressor \( s \)

• Example: Logistic-regression classifiers
Linear Separability

Linearly Separable Training Sets

- Some line (hyperplane in $\mathbb{R}^d$) separates $C, S$
- Requires *much* smaller $\mathcal{H}$
- Simplest score: $s(x) = b + w^T x$. The line is $s(x) = 0$

\[
h(x) = \begin{cases} 's' & \text{if } s(x) > 0 \\ 'c' & \text{otherwise} \end{cases}
\]
Linear separability is a property of the data \textit{in a given representation}.

- **Xform 1:** \( z = x_1^2 + x_2^2 \) implies \( x \in S \iff a \leq z \leq b \)
- **Xform 2:** \( u = |\sqrt{x_1^2 + x_2^2} - r| = |\sqrt{z} - r| \)
yields linear separability:
\[ x \in S \iff u \leq \Delta r \]