Multiclass classification

- Binary classification:
  - Number of classes = 2
  - A special case of multiclass classification

Multiclass classification
- Number of classes is > 2
Multiclass classification

- **Discriminative approach**
  - Parametric discriminant functions \( g_i(x, \Theta) \)
  - Learns discriminant functions for each class \( i \) **directly using a loss function**
    - A logistic regression model

- **Generative model approach**
  - Generative model of the distribution \( p(x, y) = p(x|y) \cdot p(y) \)
  - Learns the parameters \( \Theta \) of the models \( p(x|y) \) and \( p(y) \) using the density estimation techniques
  - Discriminant functions are based on the model
    \[
    g_i(x, \Theta) = p(y = i | x, \Theta)
    \]
    \[
    y^* = \arg \max_i g_i(x)
    \]
Multi-way classification. Example

Making class decision

**Discriminant functions:**
- **Posterior of a class** – choose the class with the highest posterior probability

**Choice:**  \( i = \arg \max_{i=0,\ldots,k-1} p(y=i \mid x, \Theta_i) \)

\[
p(y=i \mid x) = \frac{p(x \mid \Theta_i)p(y=i)}{\sum_{j=0}^{k-1} p(x \mid \Theta_j)p(y=j)}
\]
**Discriminative approach**

- **Parametric models** of discriminant functions:
  - \( g_0(x), g_1(x), \ldots, g_{K-1}(x) \)
- Learns the discriminant functions directly

**Key issues:**
- How to design the discriminant functions?
- How to train them?

**Another question:**
- Can we use binary classifiers to build the multi-class models?

---

**One versus the rest (OVR)**

**Methods based on binary classification methods**

- **Assume:** we have 3 classes labeled 0,1,2
- **Approach 1:**
  A binary logistic regression on every class versus the rest (OvR)

\[
\begin{align*}
1 & \quad 0 \text{ vs. } (1 \text{ or } 2) \\
1 & \quad 1 \text{ vs. } (0 \text{ or } 2) \\
1 & \quad 2 \text{ vs. } (0 \text{ or } 1) \\
\end{align*}
\]

**Class decision:** class label for a ‘singleton’ class
- Does not work all the time
Multiclass classification. Example

Multiclass classification. Approach 1.
**Multiclass classification. Approach 1.**

- **Ambiguous region:**
  - 0 vs. (1 or 2) classifier says 0
  - 1 vs. (0 or 2) classifier says 1

- **Region of nobody:**
  - 0 vs. (1 or 2) classifier says (1 or 2)
  - 1 vs. (0 or 2) classifier says (0 or 2)
  - 2 vs. (1 or 2) classifier says (1 or 2)

- **One solution:** compare discriminant functions defined on binary classifiers for single option:

  \[ g_i(x) = g_{i \text{ vs rest}}(w_i^T x) \]

  - discriminant function for i trained on i vs. rest

**One versus the rest (OVR)**

Unclear how to decide on class in some regions

- **Ambiguous region:**
  - 0 vs. (1 or 2) classifier says 0
  - 1 vs. (0 or 2) classifier says 1

- **Region of nobody:**
  - 0 vs. (1 or 2) classifier says (1 or 2)
  - 1 vs. (0 or 2) classifier says (0 or 2)
  - 2 vs. (1 or 2) classifier says (1 or 2)
Multiclass classification. Approach 1.

One vs One (OVO)

Methods based on binary classification methods
- **Assume**: we have 3 classes labeled 0, 1, 2
- **Approach 2**:
  - A binary logistic regression on all pairs

  \[ x_1 \rightarrow 0 \text{ vs. } 1 \]
  \[ x_i \rightarrow 0 \text{ vs. } 2 \]
  \[ x_d \rightarrow 1 \text{ vs. } 2 \]

  **Class decision**: class label based on who gets the majority
  - Does not work all the time
Multiclass classification. Example

Multiclass classification (OVO)
Multiclass classification OVO

One vs one (OVO) model

Unclear how to decide on class in some regions

- **Ambiguous region:**
  - 0 vs. 1 classifier says 0
  - 1 vs. 2 classifier says 1
  - 2 vs. 0 classifier says 2

- **One solution:** define a new discriminant function by adding the discriminant functions for pairwise classifiers

\[ g_i(x) = \sum_j (g_{i\text{ vs } j}(w^T x)) \]
Multiclass classification

OVR and OVO:
- define multiclass classifier using output classes of binary classifiers

Problems: ambiguous regions, regions of nobody
Solution: define discriminant functions for the multiclass case using the discriminant functions from binary classification problems

A Concern:
- Calibration of the discriminant functions
  - Discriminant functions from independently trained binary classification models may not be directly comparable
Solution:
- joint learning of discriminant functions