Deep Learning (CS 470, CS 570)
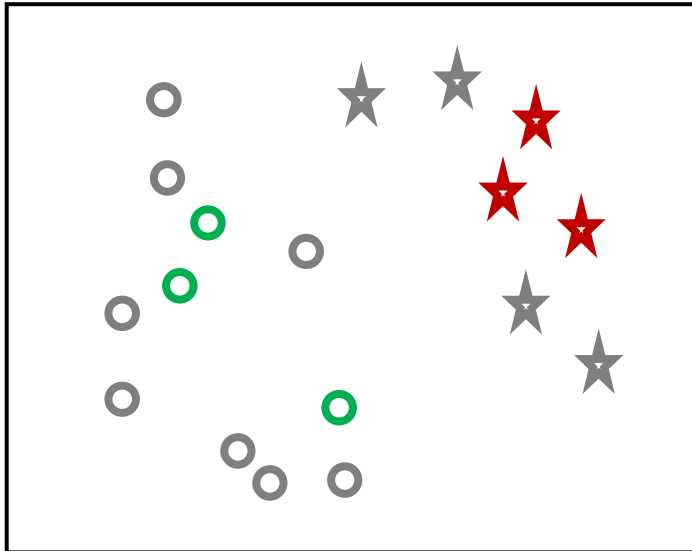
Module 2, Lecture 4: Semi-supervised Learning

Semi-supervised Learning

Some data points are labeled, while most of them are unlabeled

Why: Annotated data is costly
Needs human expertise
Unlabeled data: easily available

Data: $(\boldsymbol{x}^1, y^1), (\boldsymbol{x}^2, y^2), \ldots, (\boldsymbol{x}^M, y^M),$ $\boldsymbol{x}^{M+1}, \boldsymbol{x}^{M+2}, \ldots, \boldsymbol{x}^N$ where, $M << N$

Learn a function $\boldsymbol{h}$ such that, $h: \mathrm{x} \to y$

Algorithm:
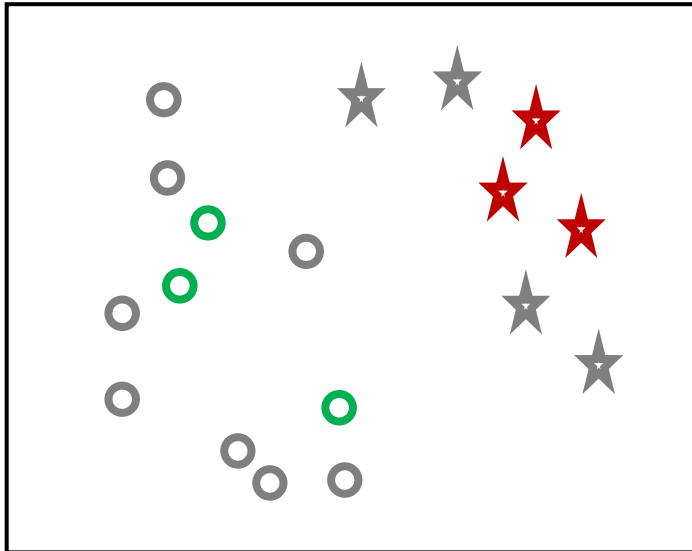
Labeled and unlabeled data

Semi-supervised Learning

Some data points are labeled, while most of them are unlabeled

Why: Annotated data is costly
Needs human expertise
Unlabeled data: easily available

Data: $(\boldsymbol{x}^1, y^1), (\boldsymbol{x}^2, y^2), \ldots, (\boldsymbol{x}^M, y^M),$ $\boldsymbol{x}^{M+1}, \boldsymbol{x}^{M+2}, \ldots, \boldsymbol{x}^N$ where, $M << N$

Learn a function $\boldsymbol{h}$ such that, $h: \mathrm{x} \rightarrow y$

Algorithm:

Labeled and unlabeled data

Predict the label of unlabeled data

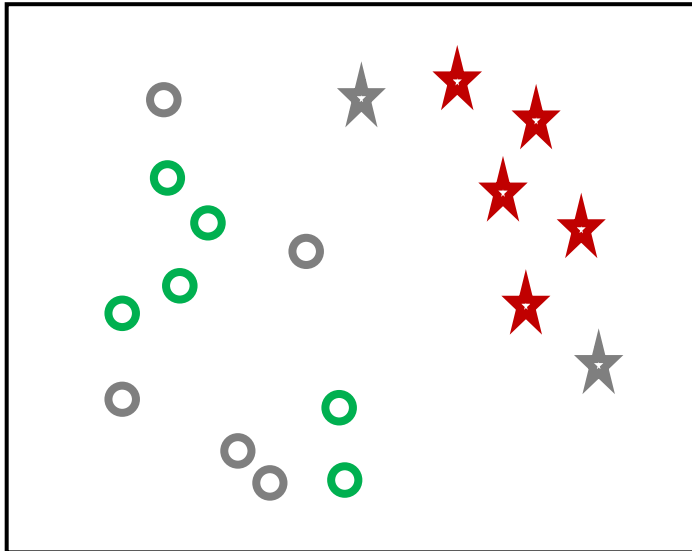Include the data predicted with higher probability in the labeled set

Semi-supervised Learning

Some data points are labeled, while most of them are unlabeled

Why: Annotated data is costly
Needs human expertise
Unlabeled data: easily available

Data: $(x^1, y^1), (x^2, y^2), \ldots, (x^M, y^M),$ $x^{M+1}, x^{M+2}, \ldots, x^N$ where, $M \ll N$

Learn a function h such that, $h: x \rightarrow y$

Algorithm:

Labeled and unlabeled data

Predict the label of unlabeled data

Include the data predicted with higher probability in the labeled set
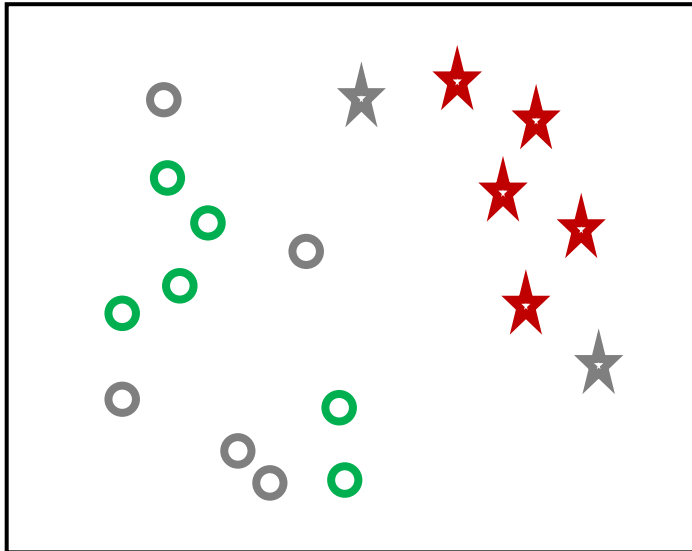
Semi-supervised Learning

Some data points are labeled, while most of them are unlabeled

Why: Annotated data is costly
Needs human expertise
Unlabeled data: easily available

Data: $(\boldsymbol{x}^1, y^1), (\boldsymbol{x}^2, y^2), \ldots, (\boldsymbol{x}^M, y^M),$ $\boldsymbol{x}^{M+1}, \boldsymbol{x}^{M+2}, \ldots, \boldsymbol{x}^N$ where, $M << N$

Learn a function $\boldsymbol{h}$ such that, $h: x \rightarrow y$

Algorithm:

Labeled and unlabeled data

Predict the label of unlabeled data

Include the data predicted with higher probability in the labeled set

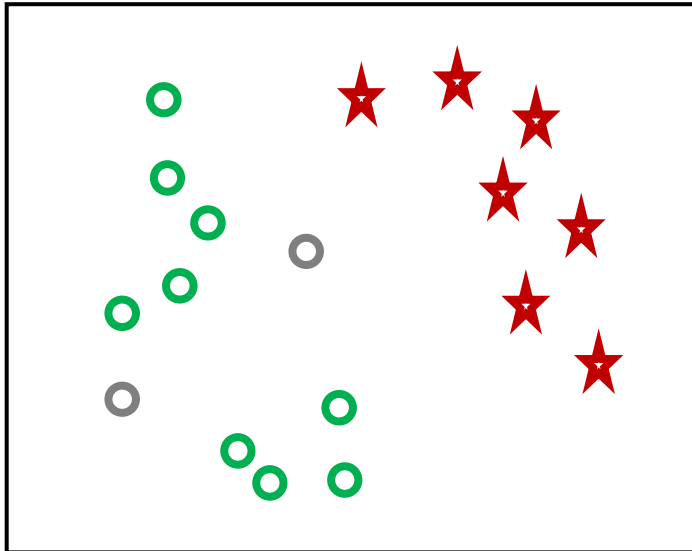Repeat it until all data points are labeled

Semi-supervised Learning

Some data points are labeled, while most of them are unlabeled

Why: Annotated data is costly
Needs human expertise
Unlabeled data: easily available

Data: $(x^1, y^1), (x^2, y^2), \ldots, (x^M, y^M),$ $x^{M+1}, x^{M+2}, \ldots, x^N$ where, $M << N$

Learn a function h such that, $h: x \to y$

Algorithm:

Labeled and unlabeled data

Predict the label of unlabeled data

Include the data predicted with higher probability in the labeled set
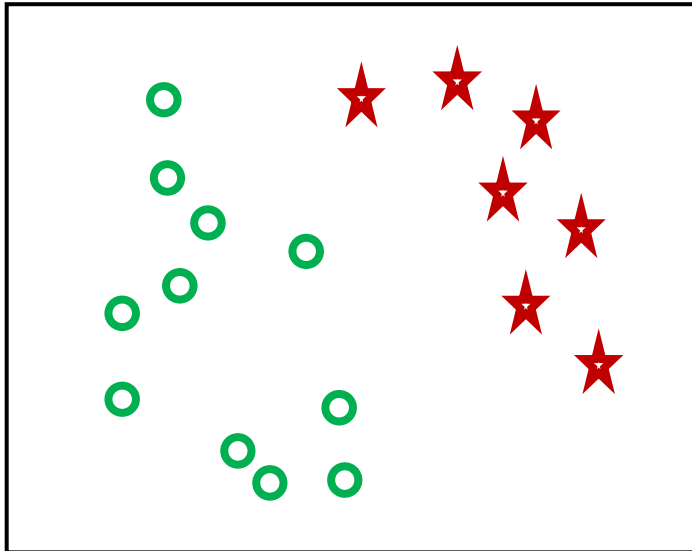
Repeat it until all data points are labeled

Semi-supervised Learning

Some data points are labeled, while most of them are unlabeled

Why: Annotated data is costly
Needs human expertise
Unlabeled data: easily available

Data: $(x^1, y^1), (x^2, y^2), \ldots, (x^M, y^M),$ $\quad x^{M+1}, x^{M+2}, \ldots, x^N$ $\quad$ where, $\quad$ M << N

Learn a function h such that, $\quad h: x \to y$

Algorithm:

Labeled and unlabeled data

Predict the label of unlabeled data

Include the data predicted with higher probability in the labeled set

Repeat it until all data points are labeled

Additional Reading

[Semi-supervised learning](#) description