Aalto University
School of Electrical Engineering

CS-C3240 – Machine Learning D

Round 3: From features to classification

Stephan Sigg

Department of Communications and Networking
Aalto University, School of Electrical Engineering
stephan.sigg@aalto.fi

Version 2.3, January 23, 2022

Overview

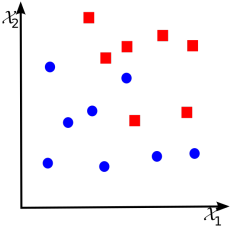
Shortage of labelled data

Shortage of labelled data

Given a set $\mathcal{Z}$ of data points $\overrightarrow{z}_1, \ldots, \overrightarrow{z}_n$, features $\mathcal{X}_1, \ldots, \mathcal{X}_m$ and labels $y_1, \ldots, y_o$, assume partially labelled sets $\overrightarrow{z}_i = \langle \overrightarrow{x}^{(i)}, y^{(i)} \rangle, i \in \{1, \ldots, l\}$ and $\overrightarrow{z}_j = \langle \overrightarrow{x}^{(j)} \rangle, j \in \{l+1, \ldots, n\}$ as well as $\overrightarrow{x}^{(k)} = x_1^{(k)}, \ldots, x_m^{(k)}, k \in \{1, \ldots, n\}$

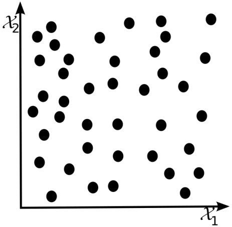
Labelled

$$\overrightarrow{x}^{(1)} \rightarrow y^{(1)}$$
$$\overrightarrow{x}^{(2)} \rightarrow y^{(2)}$$
$$\vdots$$
$$\overrightarrow{x}^{(l)} \rightarrow y^{(l)}$$



Unlabelled

$$\overrightarrow{x}^{(l+1)} \rightarrow ?$$
$$\overrightarrow{x}^{(l+2)} \rightarrow ?$$
$$\vdots$$
$$\overrightarrow{x}^{(n)} \rightarrow$$

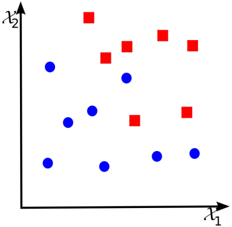


Shortage of labelled data

Given a set $\mathcal{Z}$ of data points $\vec{z}_1, \dots, \vec{z}_n$, features $\mathcal{X}_1, \dots, \mathcal{X}_m$ and labels $y_1, \dots, y_o$, assume partially labelled sets $\vec{z}_i = \langle \vec{x}^{(i)}, y^{(i)} \rangle, i \in \{1, \dots, l\}$ and $\vec{z}_j = \langle \vec{x}^{(j)} \rangle, j \in \{l+1, \dots, n\}$ as well as $\vec{x}^{(k)} = x_1^{(k)}, \dots, x_m^{(k)}, k \in \{1, \dots, n\}$

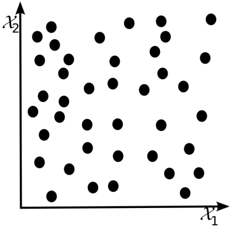
Labelled

$$\vec{x}^{(1)} \rightarrow y^{(1)}$$
$$\vec{x}^{(2)} \rightarrow y^{(2)}$$
$$\vdots$$
$$\vec{x}^{(l)} \rightarrow y^{(l)}$$



Unlabelled

$$\vec{x}^{(l+1)} \rightarrow ?$$
$$\vec{x}^{(l+2)} \rightarrow ?$$
$$\vdots$$
$$\vec{x}^{(n)} \rightarrow$$



Problem:

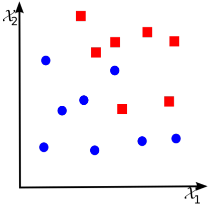
- Unlabelled training data is often easy to obtain
- **However:** labelling the data requires significant manual work

Shortage of labelled data

Automated labeling through semi-supervised learning

Increase amount of labelled data via semi-supervised learning

1. Start with labelled data $\vec{z}_i = \langle \vec{x}^{(i)}, y^{(i)} \rangle, i \in \{1, \ldots, l\}$

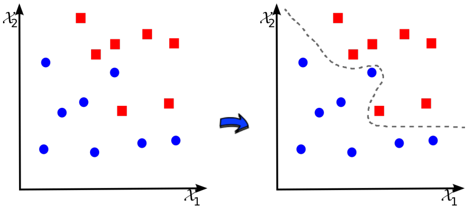


Shortage of labelled data

Automated labeling through semi-supervised learning

Increase amount of labelled data via semi-supervised learning

1. Start with labelled data $\vec{z}_i = \langle \vec{x}^{(i)}, y^{(i)} \rangle, i \in \{1, \ldots, l\}$
2. Train the classifier on the labelled data $\hat{h}\left(\hat{\vec{w}}, \cdot\right) = \min_{i=1,..l} \mathcal{L}\left(h(\vec{w}, \vec{x}^{(i)}), y^{(i)}\right)$

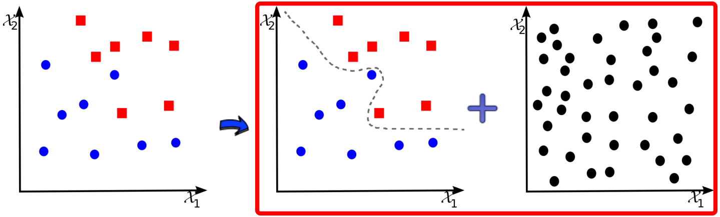


Shortage of labelled data

Automated labeling through semi-supervised learning

Increase amount of labelled data via semi-supervised learning

1. Start with labelled data $\qquad \vec{z}_i = \langle \vec{x}^{(i)}, y^{(i)} \rangle, i \in \{1, \ldots, l\}$
2. Train the classifier on the labelled data $\qquad \hat{h}\left(\hat{\vec{w}}, \cdot\right) = \min_{i=1,..l} \mathcal{L}\left(h(\vec{w}, \vec{x}^{(i)}), y^{(i)}\right)$

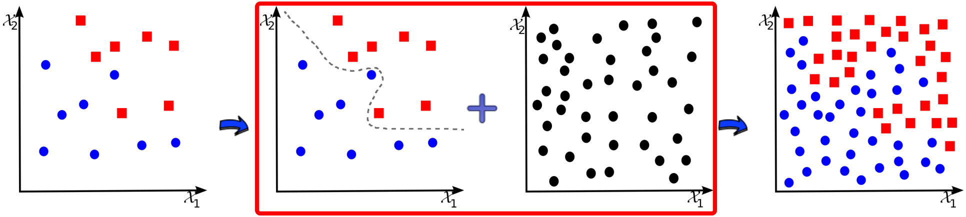


Shortage of labelled data

Automated labeling through semi-supervised learning

Increase amount of labelled data via semi-supervised learning

1. Start with labelled data $\qquad \overrightarrow{z}_i = \langle \overrightarrow{x}^{(i)}, y^{(i)} \rangle, i \in \{1, \ldots, l\}$
2. Train the classifier on the labelled data $\qquad \hat{h}\left(\overrightarrow{\hat{w}}, \cdot\right) = \min_{i=1,..l} \mathcal{L}\left(h(\overrightarrow{w}, \overrightarrow{x}^{(i)}), y^{(i)}\right)$
3. Use $\hat{h}\left(\overrightarrow{\hat{w}}, \cdot\right)$ to learn labels for $\overrightarrow{z}_j = \langle \overrightarrow{x}^{(j)} \rangle, j \in \{1, \ldots, l\}$ $\qquad \hat{y}^{(j)} = \hat{h}\left(\overrightarrow{\hat{w}}, \overrightarrow{x}^{(j)}\right)$

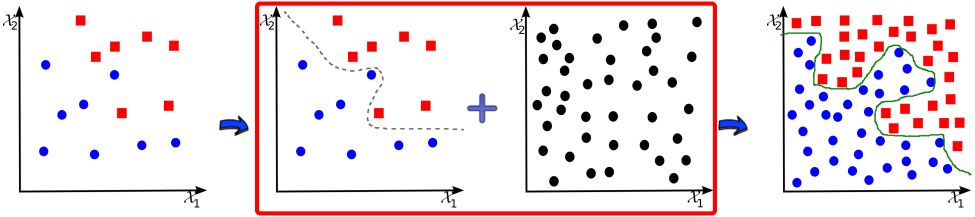


Shortage of labelled data

Automated labeling through semi-supervised learning

Increase amount of labelled data via semi-supervised learning

1. Start with labelled data $\quad \vec{z}_i = \langle \vec{x}^{(i)}, y^{(i)} \rangle, i \in \{1, \ldots, l\}$
2. Train the classifier on the labelled data $\quad \hat{h}\left(\vec{\hat{w}}, \cdot\right) = \min_{i=1,..l} \mathcal{L}\left(h(\vec{w}, \vec{x}^{(i)}), y^{(i)}\right)$
3. Use $\hat{h}\left(\vec{\hat{w}}, \cdot\right)$ to learn labels for $\vec{z}_j = \langle \vec{x}^{(j)} \rangle, j \in \{1, \ldots, l\}$ $\quad \hat{y}^{(j)} = \hat{h}\left(\vec{\hat{w}}, \vec{x}^{(j)}\right)$
4. Train new classifier $\hat{h}'$ on $\langle \vec{x}^{(1)}, y^{(1)} \rangle, \ldots, \langle \vec{x}^{(l+1)}, \hat{y}^{(l+1)} \rangle$

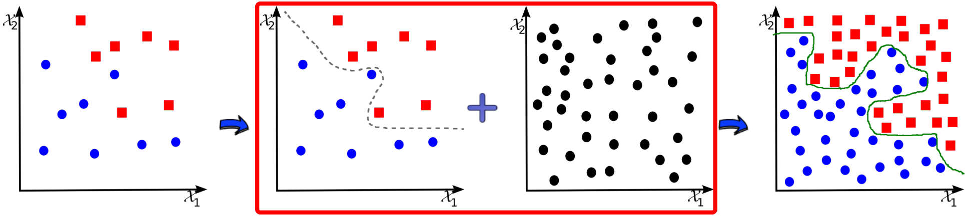


Shortage of labelled data

Automated labeling through semi-supervised learning

Remark:

- No guaranteed success → Empirical validation required
- Introducing weights to samples can reduce dependency on learned labels



Shortage of labelled data – Automatic labelling

Co-training

Provided independent feature sub-sets (perspectives) $\{\mathcal{X}\}_s$ with $\bigcup_s \{\mathcal{X}\}_s = \mathcal{X}_1, \ldots, \mathcal{X}_m$ and $\bigcap_s \{\mathcal{X}\}_s = \emptyset$, multiple classification models $h_s(\overrightarrow{w}_s, \overrightarrow{x})$ are trained to these sub-sets using the labelled data $\overrightarrow{z}_i = \langle \overrightarrow{x}^{(i)}, y^{(i)} \rangle, i \in \{1, \ldots, l\}$ to iteratively label unlabelled data $\overrightarrow{z}_j = \langle \overrightarrow{x}^{(j)} \rangle, j \in \{l+1, \ldots, n\}$

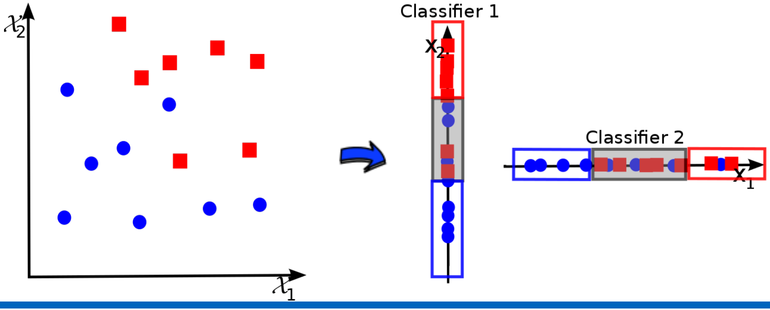
Shortage of labelled data – Automatic labelling

Co-training

Provided independent feature sub-sets (perspectives) $\{\mathcal{X}\}_s$ with $\bigcup_s \{\mathcal{X}\}_s = \mathcal{X}_1, \ldots, \mathcal{X}_m$ and $\bigcap_s \{\mathcal{X}\}_s = \emptyset$, multiple classification models $h_s(\overrightarrow{w}_s, \overrightarrow{x})$ are trained to these sub-sets using the labelled data $\overrightarrow{z}_i = \langle \overrightarrow{x}^{(i)}, y^{(i)} \rangle, i \in \{1, \ldots, l\}$ to iteratively label unlabelled data $\overrightarrow{z}_j = \langle \overrightarrow{x}^{(j)} \rangle, j \in \{l+1, \ldots, n\}$

1. Train classifier h_s for each $\{\mathcal{X}\}_s$

$$\hat{h}_s(\overrightarrow{\hat{w}_s}, \cdot) = \min_{i \in \{\mathcal{X}_s\}} \mathcal{L}\left(h(\overrightarrow{w}, \overrightarrow{x}^{(i)}), y^{(i)}\right)$$

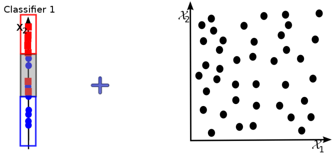


Shortage of labelled data – Automatic labelling

Co-training

Provided independent feature sub-sets (perspectives) $\{\mathcal{X}\}_s$ with $\bigcup_s \{\mathcal{X}\}_s = \mathcal{X}_1, \dots, \mathcal{X}_m$ and $\bigcap_s \{\mathcal{X}\}_s = \emptyset$, multiple classification models $h_s(\overrightarrow{w}_s, \overrightarrow{x})$ are trained to these sub-sets using the labelled data $\overrightarrow{z}_i = \langle \overrightarrow{x}^{(i)}, y^{(i)} \rangle, i \in \{1, \dots, l\}$ to iteratively label unlabelled data $\overrightarrow{z}_j = \langle \overrightarrow{x}^{(j)} \rangle, j \in \{l+1, \dots, n\}$

1. Train classifier h_s for each $\{\mathcal{X}\}_s$ $\quad \hat{h}_s(\overrightarrow{\hat{w}_s}, \cdot) = \min_{i \in \{\mathcal{X}_s\}} \mathcal{L}\left(h(\overrightarrow{w}, \overrightarrow{x}^{(i)}), y^{(i)}\right)$
2. Apply $\hat{h}_s(\overrightarrow{\hat{w}_s}, \cdot)$ to $\overrightarrow{z}_i = \langle \overrightarrow{x}^{(i)}, y^{(i)} \rangle, i \in \{1, \dots, l\}$ $\quad \hat{y}^{(j)} = \hat{h}_s(\overrightarrow{\hat{w}_s}, \overrightarrow{x}^{(j)})$

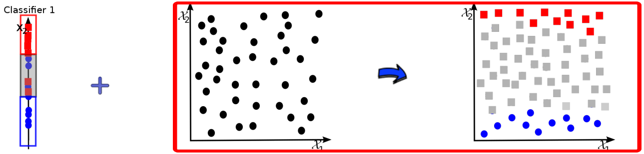


Shortage of labelled data – Automatic labelling

Co-training

Provided independent feature sub-sets (perspectives) $\{\mathcal{X}\}_s$ with $\bigcup_s \{\mathcal{X}\}_s = \mathcal{X}_1, \dots, \mathcal{X}_m$ and $\bigcap_s \{\mathcal{X}\}_s = \emptyset$, multiple classification models $h_s(\vec{w}_s, \vec{x})$ are trained to these sub-sets using the labelled data $\vec{z}_i = \langle \vec{x}^{(i)}, y^{(i)} \rangle, i \in \{1, \dots, l\}$ to iteratively label unlabelled data $\vec{z}_j = \langle \vec{x}^{(j)} \rangle, j \in \{l+1, \dots, n\}$

1. Train classifier h_s for each $\{\mathcal{X}\}_s$ — $\hat{h}_s(\vec{\hat{w}}_s, \cdot) = \min_{i \in \{\mathcal{X}_s\}} \mathcal{L}\left(h(\vec{w}, \vec{x}^{(i)}), y^{(i)}\right)$
2. Apply $\hat{h}_s(\vec{\hat{w}}_s, \cdot)$ to $\vec{z}_i = \langle \vec{x}^{(i)}, y^{(i)} \rangle, i \in \{1, \dots, l\}$ — $\hat{y}^{(j)} = \hat{h}_s(\vec{\hat{w}}_s, \vec{x}^{(j)})$
3. Add $\langle \vec{x}, \hat{y}^{(j)} \rangle$ with highest confidence to $\vec{z}_i = \langle \vec{x}^{(i)}, y^{(i)} \rangle, i \in \{1, \dots, l\}$

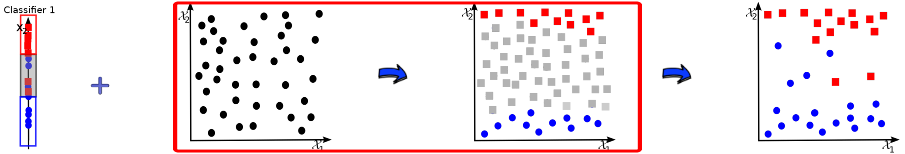


Shortage of labelled data – Automatic labelling

Co-training

Provided independent feature sub-sets (perspectives) $\{\mathcal{X}\}_s$ with $\bigcup_s \{\mathcal{X}\}_s = \mathcal{X}_1, \dots, \mathcal{X}_m$ and $\bigcap_s \{\mathcal{X}\}_s = \emptyset$, multiple classification models $h_s(\overrightarrow{w}_s, \overrightarrow{x})$ are trained to these sub-sets using the labelled data $\overrightarrow{z}_i = \langle \overrightarrow{x}^{(i)}, y^{(i)} \rangle, i \in \{1, \dots, l\}$ to iteratively label unlabelled data $\overrightarrow{z}_j = \langle \overrightarrow{x}^{(j)} \rangle, j \in \{l+1, \dots, n\}$

1. Train classifier h_s for each $\{\mathcal{X}\}_s$ $\quad \hat{h}_s(\overrightarrow{\hat{w}_s}, \cdot) = \min_{i \in \{\mathcal{X}_s\}} \mathcal{L}\left(h(\overrightarrow{w}, \overrightarrow{x}^{(i)}), y^{(i)}\right)$
2. Apply $\hat{h}_s(\overrightarrow{\hat{w}_s}, \cdot)$ to $\overrightarrow{z}_i = \langle \overrightarrow{x}^{(i)}, y^{(i)} \rangle, i \in \{1, \dots, l\}$ $\quad \hat{y}^{(j)} = \hat{h}_s(\overrightarrow{\hat{w}_s}, \overrightarrow{x}^{(j)})$
3. Add $\langle \overrightarrow{x}, \hat{y}^{(j)} \rangle$ with highest confidence to $\overrightarrow{z}_i = \langle \overrightarrow{x}^{(i)}, y^{(i)} \rangle, i \in \{1, \dots, l\}$
4. Iterate over over all classifiers h_s until convergence reached



Questions?

Stephan Sigg
stephan.sigg@aalto.fi

Si Zuo
si.zuo@aalto.fi

Literature

- C.M. Bishop: Pattern recognition and machine learning, Springer, 2007.
- R.O. Duda, P.E. Hart, D.G. Stork: Pattern Classification, Wiley, 2001.