Multi-instance learning

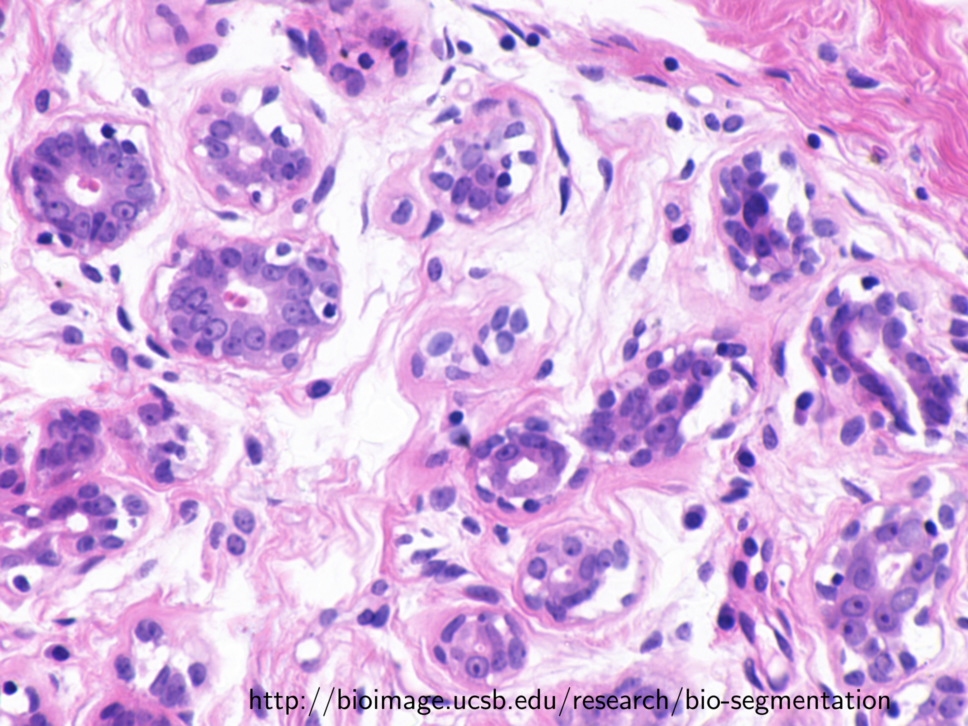
A probabilistic framework

Kajsa Møllersen

2017

MIL

- Supervised learning
- Each *bag* consists of several *instances*
- The bags are labelled, the instances are not


http://bioimage.ucsb.edu/research/bio-segmentation

Standard MI assumption

- Positive bag contains at least one positive instance
- Negative bag contains only negative instances

MI assumptions

- Both positive and negative bags contain both positive and negative instances
- Adjust assumption to problem at hand
- No-free-lunch theorem

$$X^+ : \text{random variable (vector) of positive instance}$$
$$f^+(x|\theta_b^+) : \text{pdf of positive instances in bag } b$$

$$\text{Bag } b : f_b(x) = \pi_b^+ f^+(x|\theta_b^+) + (1 - \pi_b^+) f^-(x|\theta_b^-)$$
$$0 \leq \pi_b^+ \leq 1 : \text{proportion of positive instances in bag } b$$

$$\text{Positive bag } b : f_{pos,b}(x) = \pi_{pos,b}^+ f^+(x|\theta_b^+) + (1 - \pi_{pos,b}^+) f^-(x|\theta_b^-)$$

$$\text{Negative bag } b' : f_{neg,b'}(x) = \pi_{neg,b'}^+ f^+(x|\theta_{b'}^+) + (1 - \pi_{neg,b'}^+) f^-(x|\theta_{b'}^-)$$

Both tumour and normal tissue vary from image to image:

- Assume θ_b^+ being an observation from $P(\Theta|\tau^+)$, and θ_b^- being an observation from $P(\Theta|\tau^-)$.

Tumour size varies from image to image:

- Assume $\pi_{pos,b}^+$ being an observation from $P(\Pi_{pos}^+)$.

Hierarchical distribution

Positive images

$$\begin{aligned} X_{pos}|\Theta &\sim P(X_{pos}|\Theta) \\ \Theta|\mathcal{T} &\sim P(\Theta|\mathcal{T}) \\ \mathcal{T} &\sim \begin{cases} P(\mathcal{T} = \tau^{+}) = \Pi^{+}_{pos} \\ P(\mathcal{T} = \tau^{-}) = 1 - \Pi^{+}_{pos} \end{cases} \\ \Pi^{+}_{pos} &\sim P(\Pi^{+}_{pos}) \end{aligned}$$

Similarly, for the negative images we have

$$\begin{aligned} X_{neg}|\Theta &\sim P(X_{neg}|\Theta) \\ \Theta|\mathcal{T} &\sim P(\Theta|\mathcal{T}) \\ \mathcal{T} &\sim \begin{cases} P(\mathcal{T} = \tau^{+}) = \Pi^{+}_{neg} \\ P(\mathcal{T} = \tau^{-}) = 1 - \Pi^{+}_{neg} \end{cases} \\ \Pi^{+}_{neg} &\sim P(\Pi^{+}_{neg}) \end{aligned}$$

Standard MI assumption

- $P(\Pi^+_{neg} = 0) = 1$: No positive instances in a negative bag
- $0 < \pi^+_{pos} \leq 1$: At least one positive instance in a positive bag
- $\mathcal{X}^+ \cap \mathcal{X}^- = \emptyset$: No overlap between the set of possible positive values and possible negative values

Classification of bags

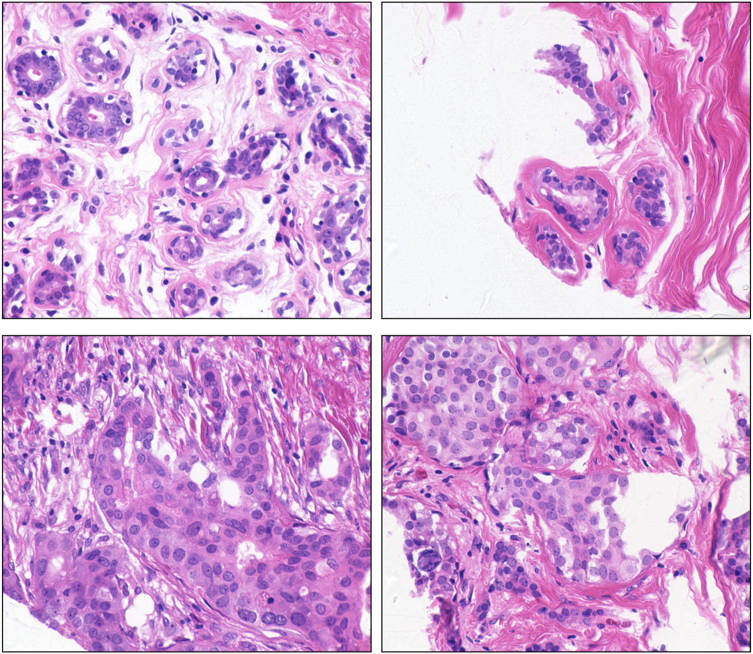
- Instance-to-instance
- Bag-to-bag
- Reduce bag to single instance. Single-instance classification

Probabilistic framework

- Describe bag as probability distribution
- Use divergence function to classify bags
- Bag-to-class

Algorithm

- Estimate class probability distributions
- Estimate (unlabelled) bag distribution
- Choose divergence function according to assumption(s)
- Classify according to penalty



Divergence

- Dissimilarity between two probability distributions
- Not necessarily mathematical distance
- No definition
- Kullback-Leibler, Bhattacharyya, Kolmogorov, Hellinger, Shannon entropy, ...

My favourite divergence

Kullback-Leibler information

$$D_{KL}(f_{bag}, f_{POS}) = \int f_{bag}(x) \log \frac{f_{bag}(x)}{f_{POS}(x)} dx.$$

Non-symmetric, f-divergence, log-ratio function.