# Active Learning

- Linear classifiers
- Support Vector Machines
- Active learning
- Cutting planes

# Linear Classifiers

- given training data $x_i \in \mathbf{R}^n$ and labels $y_i \in \{-1, +1\}^n$ for $i = 1, ..., m$
- find a linear classifier $f_{w,b}(x) = x^T w + b$ that predicts labels

# Separating Hyperplanes

- Directly estimate hyperplanes $w^T x + b \geq 0$

  parameters $\theta = (w, b)$

- hyperplane: $H = \left\{x : \; w^T x + b = 0\right\}$

- distance between a point $z$ and $H$

$$d(z, H) = \min_{h \in H} ||z - h||_2 = \frac{|w^T z + b|}{||w||_2}$$

# Margin

- margin $\rho$ of a hyperplane is defined as

$$\rho(w,b) = \min_{i=1,\ldots,n} d(x_i, H) = \min_{i=1,\ldots,n} \frac{|w^T x_i + b|}{||w||_2}$$

- maximum margin separating hyperplane is the solution of

$$\max_{w,b} \ \rho(w,b) \quad s.t. \ \ y_i(w^T x_i + b) \geq 0 \ \forall i$$

# Maximum margin hyperplane

- maximum margin separating hyperplane is the solution of

$$\max_{w,b} \min_{i=1,\ldots,n} \frac{|w^T x_i + b|}{||w||_2}$$
$$s.t.\ \ y_i(w^T x_i + b) \geq 0\ \forall i$$

- **not unique**

  $(\alpha w, \alpha b)$ corresponds to the same hyperplane as $(w, b)$ for $\alpha > 0$

- Scale $w$ and $b$ by $\frac{1}{\min_{i=1,\ldots,n} |w^T x_i + b|}$ and let $\rho = \frac{1}{||w||_2}$

# Maximum margin hyperplane classifier

$$\max_{w,b} \frac{1}{||w||_2}$$
$$s.t. \;\; y_i(w^T x_i + b) \geq 1 \; \forall i$$

equivalently

$$\min_{w,b} ||w||_2$$
$$s.t. \;\; y_i(w^T x_i + b) \geq 1 \; \forall i$$

- **hard-margin support vector machine (SVM)**

# Active Learning

- machine learning algorithms which can **actively query** a user to label new data points

- also called **optimal experimental design** in statistics

- Given a labeled dataset $x_i, y_i$, $i = 1, ..., m$, query new points $x_j$ and obtain their labels $y_j$ from an expert, for $j = 1, ...., r$

# Active Learning Strategies

- **Balance exploration and exploitation:** the choice of examples to label is seen as a dilemma between the exploration and the exploitation over the data space representation. Connected to Contextual Bandits and Thompson Sampling

- **Expected model change:** label those points that would most change the current model.

- **Expected error reduction:** label those points that would most reduce the model's generalization error.

- **Uncertainty sampling:** label those points for which the current model is least certain as to what the correct output should be.

  ⋮

# Active Learning via Cutting Planes

- Consider the set of hyperplanes that classify the labeled training data

$$\mathcal{W} := \left\{ w \,:\, y_i w^T x_i \geq 1 \, \forall i \in [m] \right\}$$

- and have large margin

$$\mathcal{M} := \left\{ w : \|w\|_2 \leq \beta \right\}$$

# Active Learning via Cutting Planes

- Apply cutting plane to the set

$$\mathcal{W} \cap \mathcal{M} = \left\{ w \, : \, y_i w^T x_i \geq 1 \, \forall i \in [m], \; \|w\|_2 \leq \beta \right\}$$

# Active Learning via Cutting Planes

- Apply cutting plane to the set

$$\mathcal{W} \cap \mathcal{M} = \left\{ w \,:\, y_i w^T x_i \geq 1 \, \forall i \in [m],\ \|w\|_2 \leq \beta \right\}$$

- Query the label of $x = x_{\text{center}}$

  where $x_{\text{center}}$ is the analytical center, center of the minimum volume ellipsoid, center of gravity of $\mathcal{W} \cap \mathcal{M}$.

- Add $(x, y)$ to the training data, update $\mathcal{W}$ and repeat

# References


[BV04] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[LR15] Ugo Louche and Liva Ralaivola. From cutting planes algorithms to compression schemes and active learning. In *2015 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2015.

[Wik] Wikipedia. Active learning. `https://en.wikipedia.org/wiki/Active_learning_(machine_learning)`.