Support vector machines unit:

- Linear programming (Wednesday)
- SVM concepts (**today**)
- Lab: SVM applications (next week Monday)
- The math of SVMs (next week Wednesday)
- SVM algorithms (next week Friday)
- (Midterm on Friday, Mar 21, after spring break)

Today:

- Comparison with other classification techniques
- Geometric intuitions
- The simplest version
- SVM hyperplanes
- Soft margin SVM
- Kernelized SVM

"The SVM provides state-of-the-art results in many applications, without sound theoretical guaantees." *Deisenroth et al, pg 335 (2020)*

"A support vector machine is a very powerful and versatile machine learning model, capable of performing linear or nonlinear classification, regression, and even outlier detection. It is one of the most popular models in machine learning, and anyone interested in machine learning should have it in their toolbox."

Geron, pg 147 (2019)

"The support vector machine is one of the most popular algorithms in modern machine learning." *Marsland, pg 169 (2015)*

"Support Vector Machines: Hype or Hallelujah?"

Title of paper by Bennett and Campbell, 2000

"Nobody uses support vector machines anymore."

ML practitioner I met a couple years ago

Given training data $\mathbf{X}, \boldsymbol{y}$, where $y_n \in \{-1, +1\}$, find $\boldsymbol{w}$, b, and r, specifically

$$\begin{aligned} \text{maximize} \quad & r \\ \text{subject to the constraints} \quad & \forall\ \boldsymbol{x_n}, y_n, \quad y_n(\boldsymbol{w}^T \boldsymbol{x_n} + b) \geq r \\ & ||\boldsymbol{w}|| = 1 \\ & r > 0 \end{aligned}$$

Equivalently, scaling so that margin $r = 1$, and then

$$\begin{aligned} \text{minmize} \quad & \tfrac{1}{2}||\boldsymbol{w}||^2 \\ \text{subject to the constraints} \quad & \forall\ \boldsymbol{x_n}, y_n, \quad y_n(\boldsymbol{w}^T \boldsymbol{x_n} + b) \geq 1 \end{aligned}$$

Given training data $\mathbf{X}, \boldsymbol{y}$, where $y_n \in \{-1, +1\}$, find $\boldsymbol{w}$, b, and r, specifically

$$\text{minimize} \quad \tfrac{1}{2}||\boldsymbol{w}||^2 + C \sum_{n=0}^{N-1} \xi_n$$

$$\text{subject to the constraints} \quad \forall\ \boldsymbol{x_n}, y_n, \quad y_n(\boldsymbol{w}^T \boldsymbol{x_n} + b) \geq 1 - \xi_n$$

. . . where ξ_n is the distance between $\boldsymbol{x}_n$ and the boundary, if $\boldsymbol{x}_n$ is misclassified; 0 otherwise.

Coming up:

Due Fri, Feb 28:
Do GMM/EM programming assignment

Due Tues, Mar 4:
Read from Chapter 5 in our textbook about SVMS
(See Canvas)

Due Wed, Mar 5:
Read and respond to Urbina et al, "Dual use of AI-powered drug discovery"
(See Canvas)

Due Fri, Mar 7:
Take SVM quiz

Due Wed, Mar 19:
Implement SVM classification

(Midterm on Fri, Mar 21)