Today

- Separability
- Generalized Linear Separatrix
- Error: Mean Squared Error

Separability:

Token sets \( X = \{x_1, \ldots, x_n\} \subseteq \mathbb{R}^d \)

and \( Z = \{z_1, \ldots, z_n\} \subseteq \mathbb{R}^d \)

are linearly separable if

\[ \exists \mathbf{w}, w_0 \text{ s.t. } \]

\[ \mathbf{w}^T \mathbf{x} + w_0 > 0 \quad \forall \mathbf{x} \in X \]

\[ \mathbf{w}^T \mathbf{z} + w_0 < 0 \quad \forall \mathbf{z} \in Z \]

A academics love:

\[ p(C_k|\mathbf{x}) = \frac{1}{1 + e^{-\mathbf{w}^T \mathbf{x} + w_0}} \]

consumers want: 100% accuracy

\[ \iff \text{perfect separability} \]
**Convex Hull Theorem**

$X$ and $Y$ are linearly separable

$H(X) \cap H(Y) = \emptyset$

$\iff$

$\exists \text{ no } x \in H(X) \quad \text{ and } \quad \exists \text{ no } y \in H(Y)$

**Convex Hull**

$H(X) = \left\{ x : \sum_{n=1}^{N} \alpha_n x^N = 0 \right\}$

$H(Y) = \left\{ y : \sum_{n=1}^{N} \beta_n y^N = 0 \right\}$

**Proof:**

$H(X) \cap H(Y) \neq \emptyset$

$\sum_{n=1}^{N} \alpha_n x_n + \sum_{n=1}^{N} \beta_n y_n = 0$

$\implies \exists x_n, y_n \geq 0$

$\text{iff } \exists x_n, y_n \geq 0$
GENERALIZED LINEAR SEPARATRIX

Suppose \( \mathbb{X}, \mathbb{Z} \) not linearly separable.

Construct nonlinear functions \( \phi_j(x) \).

Then
\[
\hat{y}(x) = \sum_{j=1}^{N} \omega_j \phi_j(x)
\]

Algorithm to guarantee generalized linear separability

1. Initialize start with \( \phi_1(x) \), \( \phi_2(x) \).

2. \( \hat{y}(x) = \sum_{j=1}^{N} \omega_j \phi_j(x) \).

II. Iterate

A. Check if convex hulls overlap

For each \( \phi_i(x) \), try to find
\[
\hat{y}^q = \sum_{k=1}^{N} \beta_k \phi_k(x)
\]
(use Lagrangian).

If no such \( \beta_k \) exist, OK.

If such \( \beta_k \) exist, \( \hat{y}^q \) is a "potential error".

B. Add an \((m+1)\)st feature

Create a \( \phi_{m+1}(x) \) that moves at least one "potential error" out of the opposing convex hull.
How to find Betas:

**Lagrangian**

\[ \mathcal{L} = \frac{1}{2} \sum_{k=1}^{n} \beta_k^2 + \frac{2}{\epsilon^2} \sum_{k=1}^{m} \lambda_k \left[ \min(0, \beta_k-1) - \max(0, \beta_k) \right] \]

\[ \mathcal{L} = \frac{1}{2} \left\| \Phi - \Psi \beta \right\|_2^2 + \frac{2}{\epsilon^2} \sum_{k=1}^{m} \lambda_k \left[ \min(0, \beta_k) - \max(0, \beta_k) \right] \]

\[ \mathcal{L} = \frac{1}{2} \left\| \Phi - \Psi \beta \right\|_2^2 + \frac{2}{\epsilon^2} \sum_{k=1}^{m} \lambda_k \left[ \min(0, \beta_k-1) - \max(0, \beta_k) \right] \]

\[ \frac{\partial \mathcal{L}}{\partial \beta_k} = \begin{cases} -\Phi^T (\Phi - \Psi \beta) & \beta_k \leq 1 \\ \Phi^T (\Phi - \Psi \beta) & \beta_k > 1 \\ \end{cases} \]

\[ \frac{\partial \mathcal{L}}{\partial \beta_k} = \begin{cases} -\Phi^T (\Phi - \Psi \beta) & \beta_k \leq 1 \\ \Phi^T (\Phi - \Psi \beta) & \beta_k > 1 \\ \end{cases} + \lambda_k \left( \begin{array}{c} 1 \\ 0 \\ \end{array} \right) \left( \begin{array}{c} \beta_k \leq 1 \\ \beta_k > 1 \end{array} \right) \]

\[ \nabla_\beta \mathcal{L} = \begin{bmatrix} \frac{\partial \mathcal{L}}{\partial \beta} & \frac{\partial \mathcal{L}}{\partial \beta_2} \\ \frac{\partial \mathcal{L}}{\partial \beta_3} & \ldots & \frac{\partial \mathcal{L}}{\partial \beta_n} \end{bmatrix} = \left( \Phi^T \Phi - \Psi^T \Psi \beta \right) + \lambda \left( \begin{array}{c} \beta_k \leq 1 \\ \beta_k > 1 \end{array} \right) \]

**Arbitrary Constant**

Choose to all of \( 0 \leq \beta_k \leq 1 \)

\[ \beta_{opt} = \max \left( 0, \min \left( m \left( \Psi^T \Psi \right)^{-1} \Psi^T \phi \right) \right) \]