Hinge Loss for Binary Classifiers

Objectives

- introduce disadvantage of squared error for classification

- introduce hinge loss cost function

- characteristics of hinge loss

Squared error "loss" can be problematic

Classifier design $\min_{w} \ell(w; \underline{A}, \underline{d}) + \lambda r(\underline{w})$ ← regularizer

loss function

squared error loss $\ell(\underline{w}; \underline{A}, \underline{d}) = \|\underline{A}\underline{w} - \underline{d}\|_2^2$

Example: dwarf planet vs. planet

object	Ceres	Eris	Pluto	Mercury	Earth	Jupiter
x_i radius ($\times 10^6$ m)	1.0	2.3	2.4	4.9	12.8	143.0
d_i label	-1	-1	-1	1	1	1

$$\underline{A} = \begin{bmatrix} 1 & 1 \\ 2.3 & 1 \\ 2.4 & 1 \\ 4.9 & 1 \\ 12.8 & 1 \\ 143 & 1 \end{bmatrix}, \quad \underline{d} = \begin{bmatrix} -1 \\ -1 \\ -1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

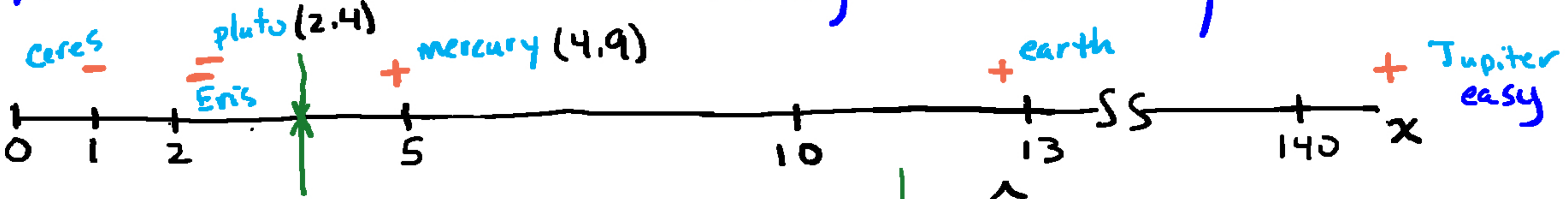
$$\underline{w}_{LS} = (\underline{A}^T\underline{A})^{-1}\underline{A}^T\underline{d} \approx 0.01 \begin{bmatrix} 1 \\ -28 \end{bmatrix}$$

dwarf: $x_i < 28$ (earth!)

planet: $x_i > 28$

squared error → poor classification

Avoid loss due to "easy-to-classify" data



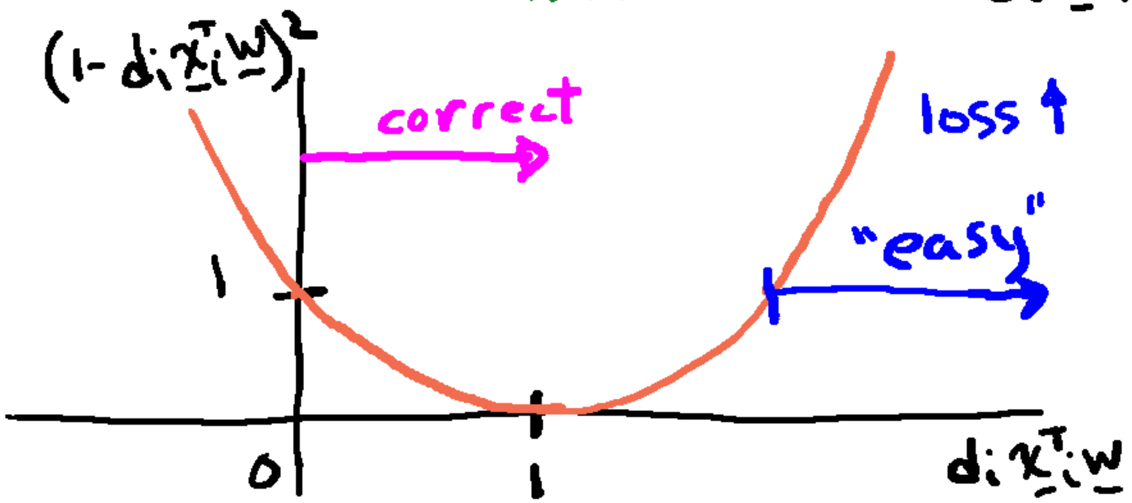
max margin classifier: midpoint $\hat{d} = \text{sign}(x - 3.65)$

Margin $4.9 - 2.4 = 2.5$ (class separation)

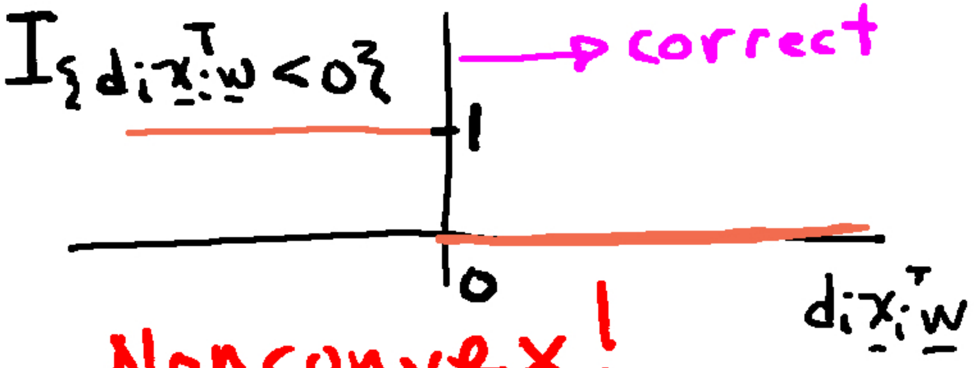
Squared error loss: $\|\underline{A}\,\underline{w} - \underline{d}\|_2^2 = \sum_{i=1}^{N} (d_i - \underline{x}_i^T \underline{w})^2 = \sum_i (1 - d_i \underline{x}_i^T \underline{w})^2$

$(d_i = \pm 1)$

correct classification: $d_i \underline{x}_i^T \underline{w} > 0$



Ideal loss

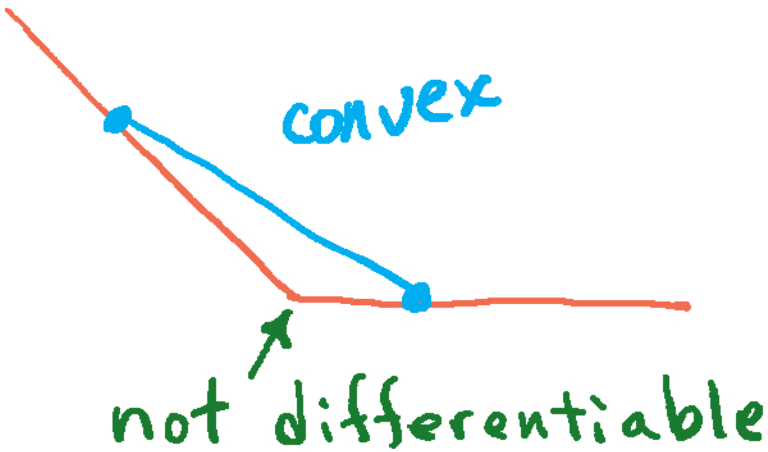


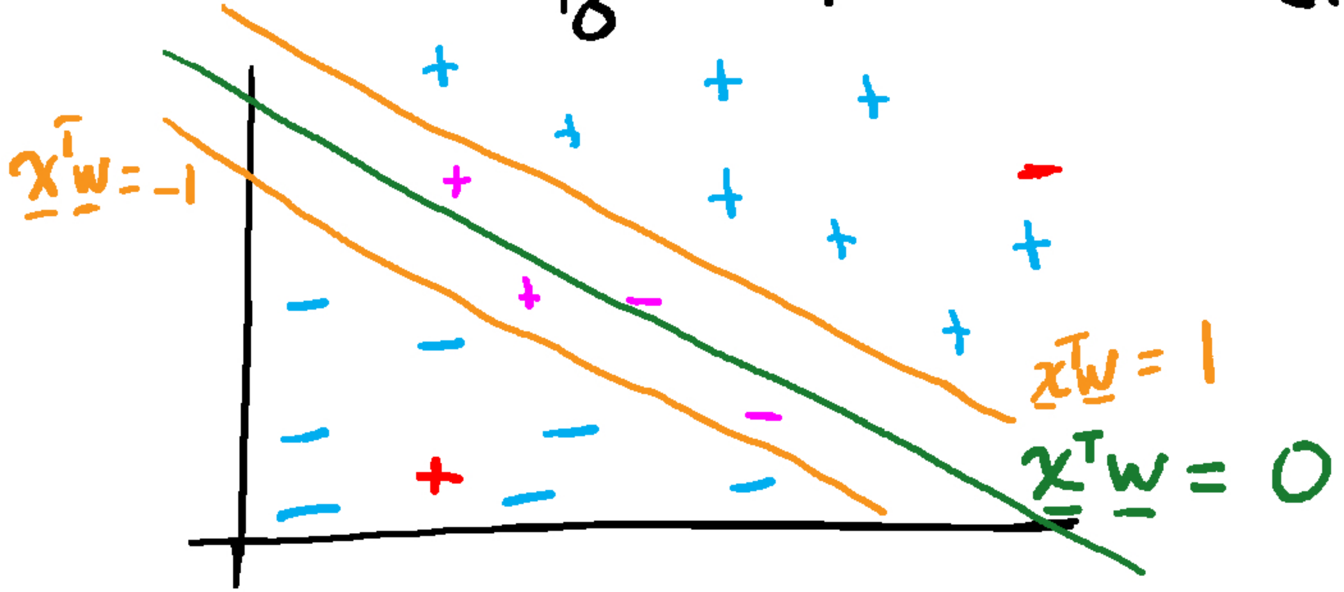
Nonconvex!

Hinge loss is convex and has no loss for easy to classify data

$$\ell(\underline{w}; \underline{A}, \underline{d}) = \sum_{i=1}^{N} (1 - d_i \underline{x}_i^T \underline{w})_+$$

$$(\alpha)_+ = \begin{cases} \alpha & \alpha > 0 \\ 0 & \alpha \le 0 \end{cases}$$





+, − no hinge loss

+, − small hinge loss

+, − large hinge loss

Hinge loss better approximates ideal: number of misclassifications

Iterative algorithms required for finding minimum hinge loss classifier