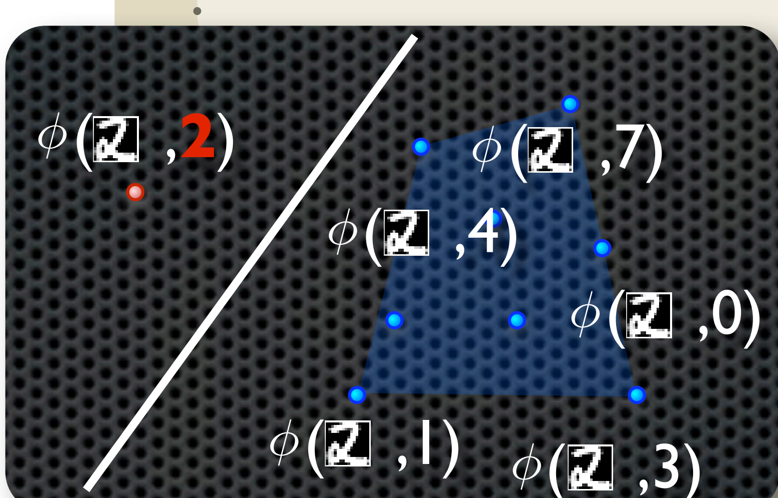
Structured SVM

Structured Prediction

Goal: Learn mapping from inputs $\boldsymbol{x} \in \mathcal{X}$
to structured outputs $\boldsymbol{y} \in \mathcal{Y}$.

SVM: Given a joint "structured" feature map $\phi : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}^d$, construct a good classifier of the form

$$h_{\boldsymbol{w}}(\boldsymbol{x}) = \underset{\boldsymbol{y} \in \mathcal{Y}}{\operatorname{argmax}} \langle \boldsymbol{w}, \boldsymbol{\phi}(\boldsymbol{x}, \boldsymbol{y}) \rangle$$



Large margin separation

Maximization oracle
(loss augmented decoding)

Primal

$$\min_{\boldsymbol{w}} \quad \frac{\lambda}{2} \|\boldsymbol{w}\|^2 + \frac{1}{n} \sum_{i=1}^{n} \underbrace{\max_{\boldsymbol{y} \in \mathcal{Y}} \Big\{ L(\boldsymbol{y}_i, \boldsymbol{y}) - \langle \boldsymbol{w}, \underbrace{\boldsymbol{\phi}(\boldsymbol{x}_i, \boldsymbol{y}_i) - \boldsymbol{\phi}(\boldsymbol{x}_i, \boldsymbol{y})}_{=: \boldsymbol{\psi}_i(\boldsymbol{y})} \rangle \Big\}}_{\text{= structured hinge loss}}$$

primal-dual correspondence
$\boldsymbol{w} = A\boldsymbol{\alpha}$

Dual

$$\min_{\boldsymbol{\alpha} \in \mathbb{R}^{n \cdot |\mathcal{Y}|}} \quad f(\boldsymbol{\alpha}) := \frac{\lambda}{2} \|A\boldsymbol{\alpha}\|^2 - \boldsymbol{b}^T \boldsymbol{\alpha}$$

$$\text{s.t.} \quad \textstyle\sum_{\boldsymbol{y} \in \mathcal{Y}} \alpha_i(\boldsymbol{y}) = 1 \quad \forall i \in [n]$$

$$\text{and} \quad \alpha_i(\boldsymbol{y}) \geq 0 \quad \forall i \in [n], \forall \boldsymbol{y} \in \mathcal{Y}$$

block-structure!

Challenge: exponential # of dual variables

$A := \left\{ \frac{1}{\lambda n} \psi_i(\boldsymbol{y}) \in \mathbb{R}^d \;\middle|\; i \in [n], \boldsymbol{y} \in \mathcal{Y} \right\}$ $\qquad$ $\boldsymbol{b} := \left(\frac{1}{n} L_i(\boldsymbol{y}) \right)_{i \in [n], \boldsymbol{y} \in \mathcal{Y}}$