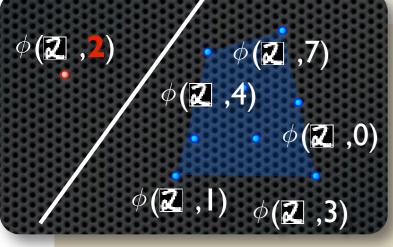
Structured SVM

Structured Prediction

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

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

 $\min_{\boldsymbol{w}} \quad \frac{\lambda}{2} \|\boldsymbol{w}\|^2 + \frac{1}{2} \sum_{\boldsymbol{m}}^{n} \boldsymbol{m}$

Maximization oracle (loss augmented decoding)

 $+\frac{1}{n}\sum_{i=1}^{n}\max_{\boldsymbol{y}\in\mathcal{Y}}\left\{L(\boldsymbol{y}_{i},\boldsymbol{y})-\left\langle\boldsymbol{w},\boldsymbol{\phi}(\boldsymbol{x}_{i},\boldsymbol{y}_{i})-\boldsymbol{\phi}(\boldsymbol{x}_{i},\boldsymbol{y})\right\rangle\right\}$

= structured hinge loss $=: \psi_i(y)$

Jual

 $\min_{oldsymbol{lpha} \in \mathbb{R}^{n \cdot |\mathcal{Y}|} \quad f(oldsymbol{lpha}) \ := \ rac{\lambda}{2} \left\| A oldsymbol{lpha}
ight\|^2 - oldsymbol{b}^T oldsymbol{lpha}$

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

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

Challenge: exponential # of dual variables

 $A := \left\{ \frac{1}{\lambda n} \boldsymbol{\psi}_i(\boldsymbol{y}) \in \mathbb{R}^d \mid i \in [n], \boldsymbol{y} \in \mathcal{Y} \right\}$

 $oldsymbol{b} := \left(rac{1}{n}L_i(oldsymbol{y})
ight)_{i \in [n], oldsymbol{y} \in \mathcal{Y}}$

primal-dual correspondence

 $\mathbf{w} = A\boldsymbol{\alpha}$

block-structure!