STRUCTURED PREDICTION

$\mathcal{Y}$ has complex struc. other than $\mathbb{R}$.

→ Cumbersome to (linearly) parameterize $g : \mathcal{X} \to \mathcal{Y}$.

→ Alternative: pose as regression problem

find $g : \underline{\mathcal{X}} \to \mathbb{R}$, where $\underline{\mathcal{X}} = \mathcal{X} \times \mathcal{Y}$

to learn compatibility of pairs (x, y).

(i) → How to get training set?

(ii) → How to get kernel over $\mathcal{X} \times \mathcal{Y}$

→ k_1 is kernel over $\mathcal{X}$, k_2 is kernel over $\mathcal{Y}$,

Any analytic function (≥ 0 coeff.) is a kernel over $\mathcal{X} \times \mathcal{Y}$

(iii) → How to infer label given x using 'g'.

→ find most compatible:

$$\hat{y} = \underset{y \in \mathcal{Y}}{\operatorname{argmax}} \; g(x, y)$$

Answer (i), insist on,

$$w^T \varphi(x_i, y_i) - \max_{y \in \mathcal{Y} \setminus \{y_i\}} w^T \varphi(x_i, y) \geq 0$$

Again, ||ly situation of one free parameter.

→ Maximize margin,

$$w^T \varphi(x_i, y_i) - \max_{y \in \mathcal{Y} \setminus \{y_i\}} w^T \varphi(x_i, y) \geq 1$$

→ Maximize margin, minimize loss (hinge?)

$$\min_{w \in \mathbb{R}^n} \frac{1}{2} \|w\|^2 + \frac{C}{m} \sum_{i=1}^{m} \xi_i$$

s.t.

$$\left. \begin{aligned} w^T \varphi(x_i, y_i) - \max_{y \in \mathcal{Y} \setminus \{y_i\}} w^T \varphi(x_i, y) &\geq 1 - \xi_i, \\ \xi_i &\geq 0 \end{aligned} \right\} \forall i$$

$$l\left(y_i, g(x_i)\right) \equiv \max\left(0,\ 1-\left(w^T\varphi(x_i, y_i) - \max_{y \in Y \setminus \{y_i\}} w^T\varphi(x_i, y)\right)\right)$$