BTP

Paper: Tomasso Poggio & Steve Smale.

(1) $S_m = (x_i, y_i)_{i=1}^{m}$

(2) Check $K_x(x') = K(x, x')$ — symmetric +ve definite f^n

Ex: $K(x, x') = e^{-\|x-x'\|^2/2\sigma^2}$.

(3) Define $f(x) = \sum_{i=1}^{m} c_i K_{x_i}(x)$ and

(4) $(m\gamma I + K)c = y$.

→ eventually 3 variables needed γ, c, K.

γ → +ve real no.; c → m dim. vector; K → kernel

Proof: $\min_{f} \left[\frac{1}{m} \sum_{i=1}^{m} (f(x_i) - y_i)^2 \right]$ → ill posed (since choice of hypothesis H).

So we minimize f over H_K.

$$\min_{f,\ f \in H_K} \frac{1}{m} \sum_{i=1}^{m} (y_i - f(x_i))^2 + \gamma \|f\|_K^2.$$

↳ ~~norm~~ norm in H_K. defined by kernel K.

↳ Defining the norm???

Differentiating above: $\frac{1}{m} \sum_{i=1}^{m} (y_i - f(x_i)) \bar{f}(x_i) - \gamma \langle f, \bar{f} \rangle = 0$

Put $\bar{f} = K_x$.

$f(x) = \sum_{i=1}^{m} c_i K_{x_i}(x)$, where $c_i = \frac{y_i - f(x_i)}{m\gamma}$.

BTP

Date / DELTA Pg No.

Sir's paper

Data $g_0, g_{-1}, \dots g_{-m+1}$ at $t = t_0$.

$\underbrace{g_0, g_{-1}, \dots g_{-m+1}}_{m\text{-preceding readings}}$ $SH = t_0 - t_{-m+1}$

Predict $g = g(t)$ for $\{t_j\}_{j=1}^{n}$ $PH = t_n - t_0$.

Training Set: $z = \{(x_\mu, y_\mu), \mu = 1, 2, \dots, M\}, |z| = M$.

$x_\mu = ((t^\mu_{-m+1}, g^\mu_{-m+1}), \dots, (t^\mu_0, g^\mu_0)) \in (\mathbb{R}^2_+)^m$

$y^\mu = ((t^\mu_1, g^\mu_1) \dots (t^\mu_n, g^\mu_n)) \in (\mathbb{R}^2_+)^n$.

Prediction func^n: $f_z : (\mathbb{R}^2)^m \to (\mathbb{R}^2)^n$

$$\min_{\substack{f_z \\ f_z \in H_K}} \frac{1}{|z|} \sum_{\mu=1}^{|z|} \|f(x_\mu) - y_\mu\|^2_{(\mathbb{R}^2)^n} + \lambda \|f\|^2_{H_K}.$$

f is RKHS $\Rightarrow$ $f(x) = K_x^* f$, $K_x^* : H \to (\mathbb{R}^2)^n$

- 3 Phases of meta-learning

(1) Choosing $K = K^\mu$ & $\lambda = \lambda_\mu$ for each μ.

(2) $\{u_\mu\}$ → Heuristically taken. meta-features

(3) Learning at meta-level. Relationship b/w $\{u_\mu\}$ & λ^μ.

- Optimization operation — Performed both (1) & (3) places

(1) $(x_\mu, y_\mu) \in (\mathbb{R}^2_+)^m \times (\mathbb{R}^2_+)^n$.

$\Rightarrow$ ↳ rephrase it to scalar.

θ $U \to V$.

$\mathbb{R}^d$ $\mathbb{R}$

(2) Split 2 sets W_1 & W_2 $\subset U \times V$.

(3) $f_\lambda = f_\lambda(\cdot\,; K, W_1) = \arg\min T_\lambda(f; K, W_1)$ ↳ (6). $\Big\}$ W_1

~~↳ which then comes out~~ $f_\lambda = f_{\lambda(K)} = \sum_i c_i^\lambda K(\cdot, u_i)$

$P(f_\lambda; W_2) = \frac{1}{|W_2|} \sum p(f_\lambda(u_i), v_i)$ $\Big\}$ W_2.

(4) $\min\ Q_\theta(K, \lambda, W_1, W_2) = \theta\, T_\lambda(f_\lambda(\cdot\,; K, W_1); K, W_1)$.

$+ (1-\theta)\, P(f_\lambda(\cdot\,; K, W_1); W_2)$

- Example:

(1) Use opti to find K^μ & λ_μ for each μ. → See how opt. is applied.

(2) $z_i = \{(u_\mu, \omega^\theta_{i,\mu}), \mu = 1, 2, \dots, M\}$, $i = 1, 2, 3 \dots$.

→ This z_i covers initial z data.

(3) Define U & V for opt. again. ↳ (15) $U \to u_\mu \in \mathbb{R}^2$.

$V \to \omega^\circ_{i,\mu} \in \mathbb{R}$.

and λ°_μ.