



دانشگاه صنعتی امیرکبیر، دانشکده ی مهندسی برق



بررسی و پیاده سازی الگوریتم های بیزوی (Bayesian)
در مدل سازی سیگنال های چندبعدی
دارای همبستگی بین ابعادی

Multitask Learning

دانیال خشابی

تیر ۱۳۹۱

مطالب ارائه شده

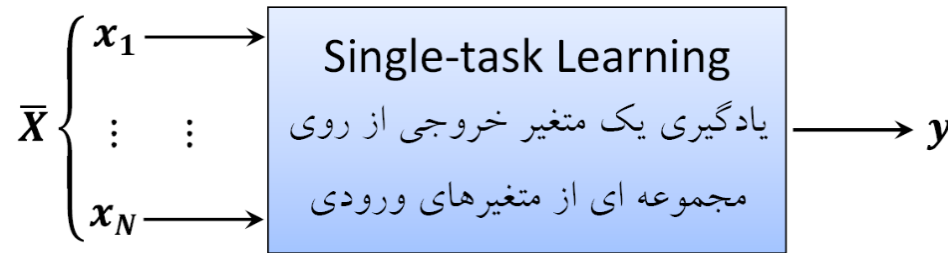


- تعریف مساله و اهمیت آن
- روش های یادگیری بیزوی بدون همبستگی متغیرها
- روش های یادگیری بیزوی با همبستگی متغیرها
- نتایج شبیه سازی و نتیجه گیری

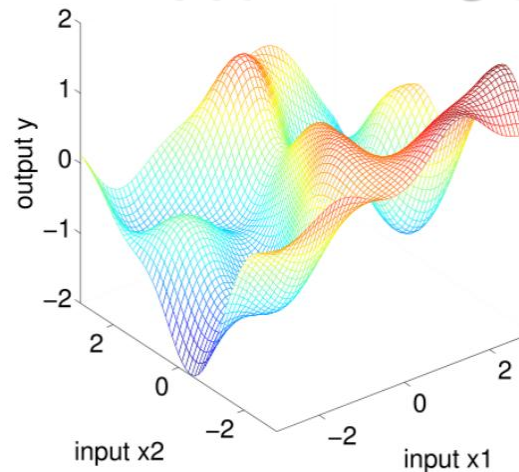
تعریف مساله (۱)



- یادگیری با تک خروجی:



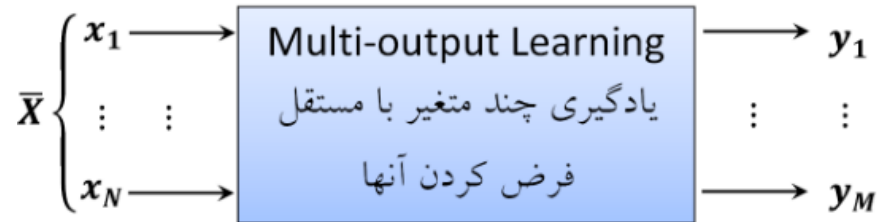
- مثل: هر نوع رگرسیون با خروجی یک بعد (تصویر از [1])



تعریف مساله (۲)

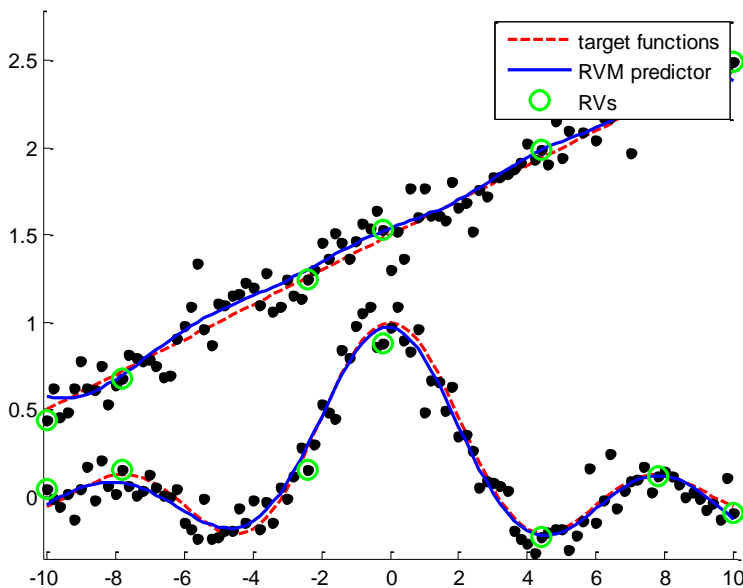


• یادگیری با چند خروجی:



• مثل:

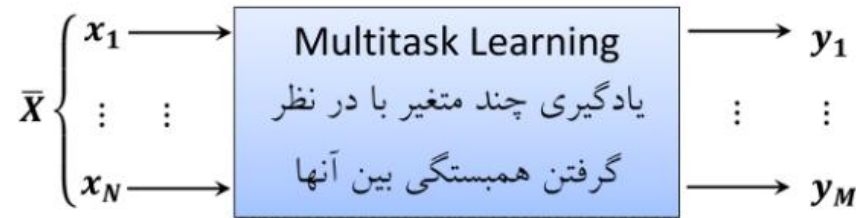
- هر نوع رگرسیون های مستقل برای چند خروجی
- دینامیک معکوس روبات
- (نگاشت ۲۱ متغیر ورودی به ۷ گشتاور موتور)
- نگاشت فضای ویژگی ها به چند کلاس
- در تشخیص احساسات فرد



تعریف مساله (۳)



• یادگیری با خروجی های همبسته (Multitask Learning)

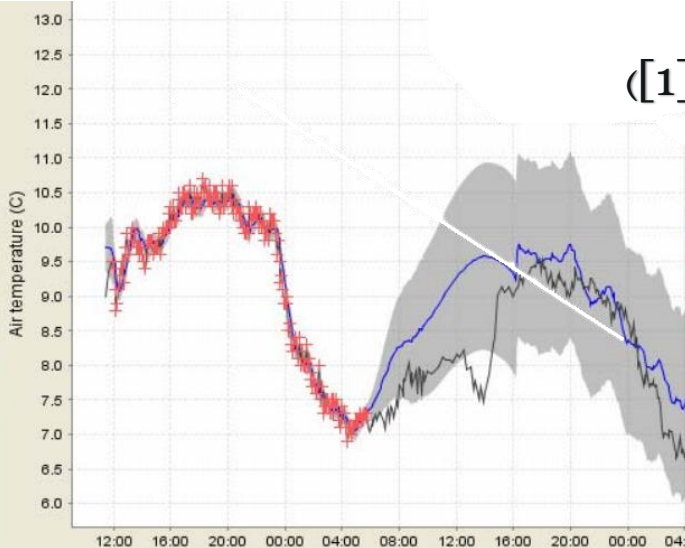


• مثال:

• شبکه های حسگر: دارای داده های نمونه برداری شده با همبستگی بالا

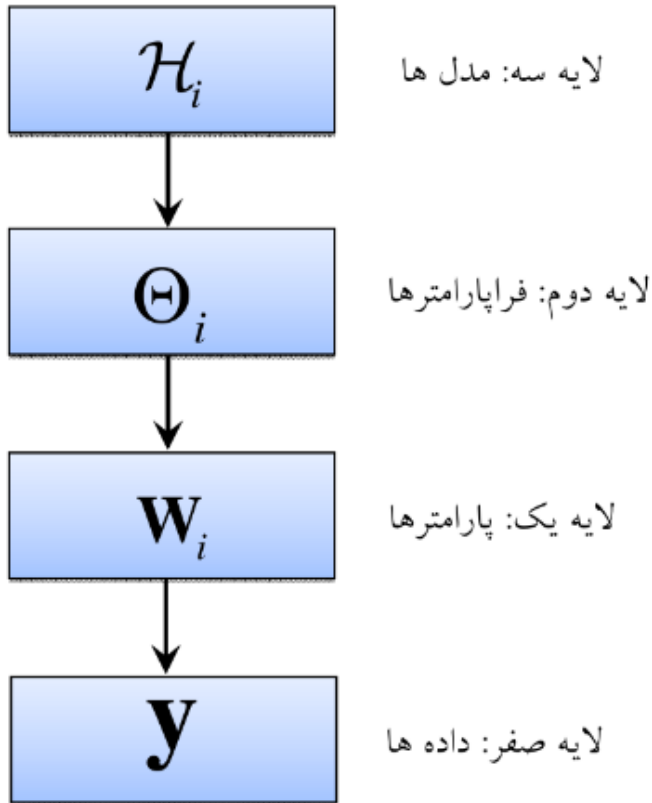
• داده های مربوط به توزیع فلزات مختلف در یک منطقه

• داده های هواشناسی مربوط به چند شهر بسیار نزدیک (تصویر از [1])



[1] Osborne et al, Towards real-time information processing of sensor network data using computationally efficient multi-output gaussian processes. IPSN '08, pp. 109–120.

تعریف مساله (۴)



چرا یادگیری بیزوی (Bayesian)؟

- قدرت مدل سازی بسیار بالا
- استنتاج احتمالی

اشکالات:

- آموزش بسیار دشوار مدل
- روش های آموزش متداول:

Variational EM

Automatic Density Filtering

Expectation Propagation

Markov Chain Monte Carlo (MCMC)

....

روش های یادگیری پیروی (۱)



روش Gaussian Process

- با فرض $\mathbf{X} = [\mathbf{x}_1 \dots \mathbf{x}_n]^T$ و $\mathbf{f} = [f_1 \dots f_n]^T$ مجموعه ای از متغیرهای تصادفی تعریف شده در ورودی، که دارای توزیع مشترک گوسی روی نقاط متناظرشان در خروجی هستند.

$$p(\mathbf{f}|\mathbf{X}) \sim \mathcal{N}(0, \mathbf{K}_{\mathbf{f},\mathbf{f}})$$

$$K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j) \quad \text{ماتریس کواریانس:}$$

$$k(\mathbf{x}_i, \mathbf{x}_j) = \sigma_f^2 \exp\left[-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2l^2}\right] \quad \text{• به عنوان مثال:}$$

- آموزش: با داشتن داده های جدید $\mathbf{X}^* = [\mathbf{x}_1^*, \dots, \mathbf{x}_m^*]^T$ و $\mathbf{f}^* = [f_1^*, \dots, f_m^*]^T$

$$p(\mathbf{f}, \mathbf{f}^*|\mathbf{X}) \sim \mathcal{N}\left(0, \begin{bmatrix} \mathbf{K}_{\mathbf{f},\mathbf{f}} & \mathbf{K}_{*,\mathbf{f}} \\ \mathbf{K}_{\mathbf{f},*} & \mathbf{K}_{*,*} \end{bmatrix}\right)$$

$$p(\mathbf{f}^*|\mathbf{f}, \mathbf{X}) = \mathcal{N}\left(\mathbf{K}_{*,\mathbf{f}}\mathbf{K}_{\mathbf{f},\mathbf{f}}^{-1}\mathbf{y}, \mathbf{K}_{*,*} - \mathbf{K}_{*,\mathbf{f}}\mathbf{K}_{\mathbf{f},\mathbf{f}}^{-1}\mathbf{K}_{\mathbf{f},*}\right)$$

- اشکال: حجم محاسبات بالا، پیچیدگی محاسبات $O(n^3)$

روش های یادگیری بیزی (۲)



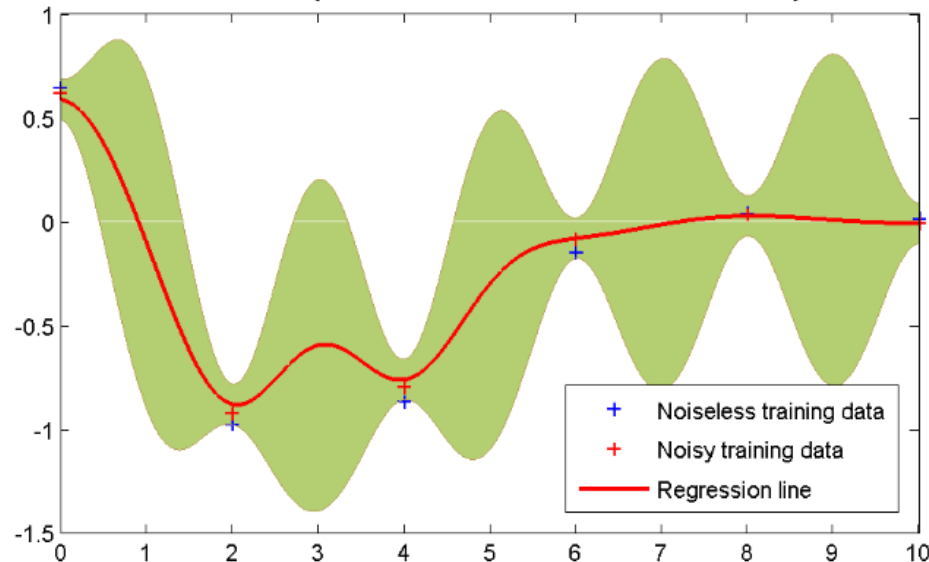
روش Gaussian Process

- آموزش مدل: با استفاده از گرادیان درست نمایی حاشیه ای نسبت به پارامترها

$$\frac{\partial}{\partial \theta_j} \log p(\mathbf{y}|\mathbf{X}, \Theta) = -\frac{1}{2} \mathbf{y}^T \mathbf{K}_y^{-1} \frac{\partial \mathbf{K}_y}{\partial \theta_j} \mathbf{K}_y^{-1} \mathbf{y} - \frac{1}{2} \text{tr} \left(\mathbf{K}^{-1} \frac{\partial \mathbf{K}}{\partial \theta_j} \right)$$

$$k(\mathbf{x}_i, \mathbf{x}_j) = \sigma_f^2 \exp \left\{ -\frac{1}{2} (\mathbf{x}_i - \mathbf{x}_j)^T M (\mathbf{x}_i - \mathbf{x}_j) \right\} + \sigma^2 \delta_{i,j}.$$

- مثال:

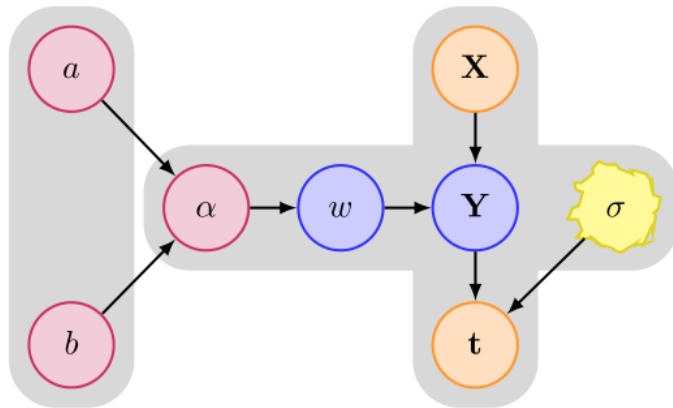


روش های یادگیری پیروی (۳)



روش های ارائه شده برای ساده سازی عملیات در GP:

• مدل Relevance Vector Machine :



$$\mathbf{t} = \Phi \mathbf{w} + \epsilon$$

$$\epsilon \sim \mathcal{N}(0, \sigma^2)$$

$$w_i \sim \mathcal{N}(0, \alpha_i^{-1})$$

• آموزش مدل: EM

• مدل Informative Vector Machine

• ایده: همه ی داده های آموزشی دارای اطلاعات یکسان نیستند!

• آموزش مدل: Automatic Density Filtering

• روش های جبری و عددی برای ساده سازی عکس ماتریس

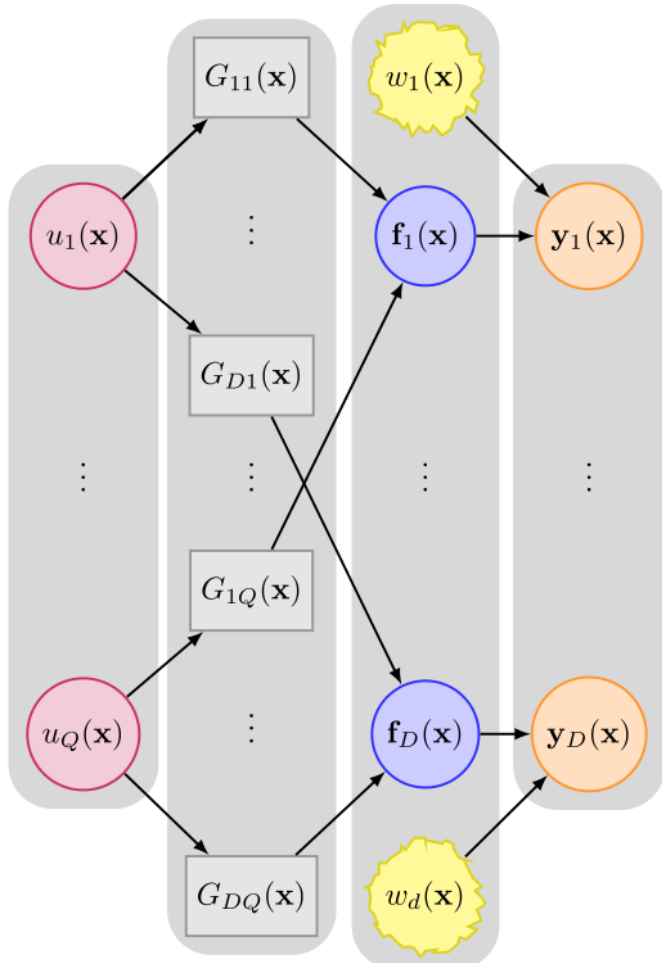
• مثل روش های:

• Subset of Regressors (SoR)

• Deterministic Training Conditional (DTC)

• و

یادگیری پیروی با همبستگی بین متغیرها (۱)



• مدل کانولوشنی (ارائه شده در [1]):

• مدل:

$$y_d(\mathbf{x}) = f_d(\mathbf{x}) + w_d(\mathbf{x})$$

$$w_d(\mathbf{x}) \sim \mathcal{N}(\mathbf{x} | \mathbf{0}, \sigma^2)$$

$$G_{d,q} = S_{d,q} \mathcal{N}(\mathbf{x} | \mathbf{0}, \mathbf{P}_d^{-1})$$

$$k_q(\mathbf{x}, \mathbf{x}') = \mathcal{N}(\mathbf{x} - \mathbf{x}' | \mathbf{0}, \mathbf{\Lambda}_q^{-1})$$

$$p(\mathbf{y} | \mathbf{X}, \theta) = \mathcal{N}(\mathbf{0}, \mathbf{K}_{f,f} + \sigma \mathbf{I})$$

• توزیع پیش بینی:

$$p(\mathbf{y}^* | \mathbf{y}, \mathbf{X}^*, \mathbf{X}, \Theta) = \mathcal{N}(\mathbf{y}^* | \mathbf{K}_{*,f} \mathbf{K}_{f,f}^{-1} \mathbf{y}, \mathbf{K}_{*,f} \mathbf{K}_{f,f}^{-1} \mathbf{K}_{f,*})$$

• آموزش پارامترها: مشابه GP

یادگیری پیروی با همبستگی بین متغیرها (۲)



روش Spike and Slab (ارائه شده در [1])

• مدل: با فرض داده های آموزشی $\mathcal{D} = \{\mathbf{X}, \mathbf{Y}\} = \{\mathbf{x}_n, \mathbf{y}_n\}_{n=1}^N$

که در آن $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N] \in \mathbb{R}^{N \times D}$ و $\mathbf{Y} \in \mathbb{R}^{N \times Q}$

$$y_{nq} \sim \mathcal{N}(y_{nq} | f_q(\mathbf{x}_n), \sigma_q^2)$$

$$f_q(\mathbf{x}) = \sum_{m=1}^M w_{qm} \phi_m(\mathbf{x}) = \mathbf{w}_q^T \phi(\mathbf{x})$$

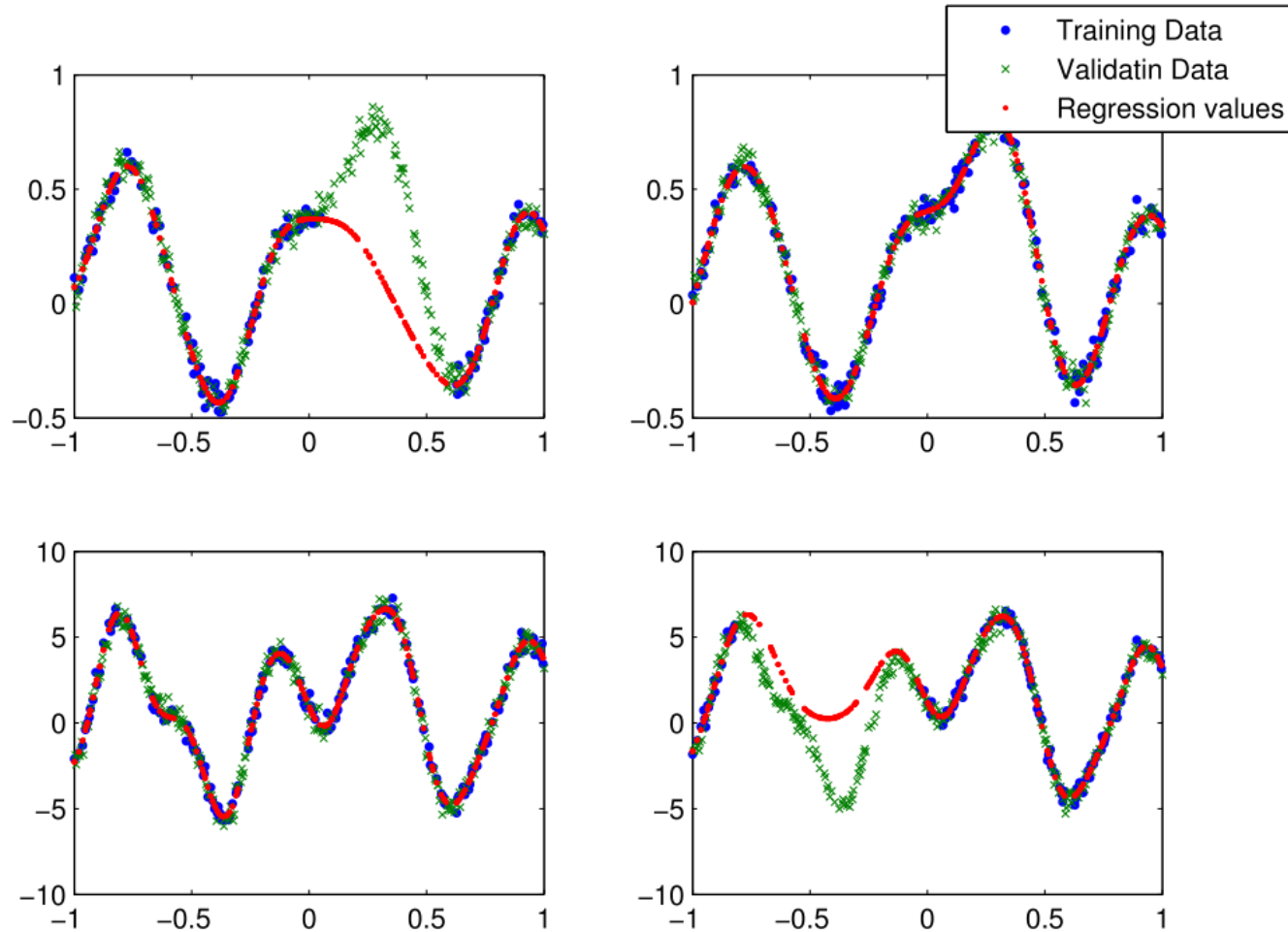
$$w_{qm} \sim \pi \mathcal{N}(w_{qm} | 0, \sigma_w^2) + (1 - \pi) \delta_0(w_{qm})$$

$$\phi_m(\mathbf{x}) \sim \mathcal{GP}(\mu_m(\mathbf{x}), k_m(\mathbf{x}_i, \mathbf{x}_j))$$

• آموزش مدل: روش Variational EM

نتایج (۱)

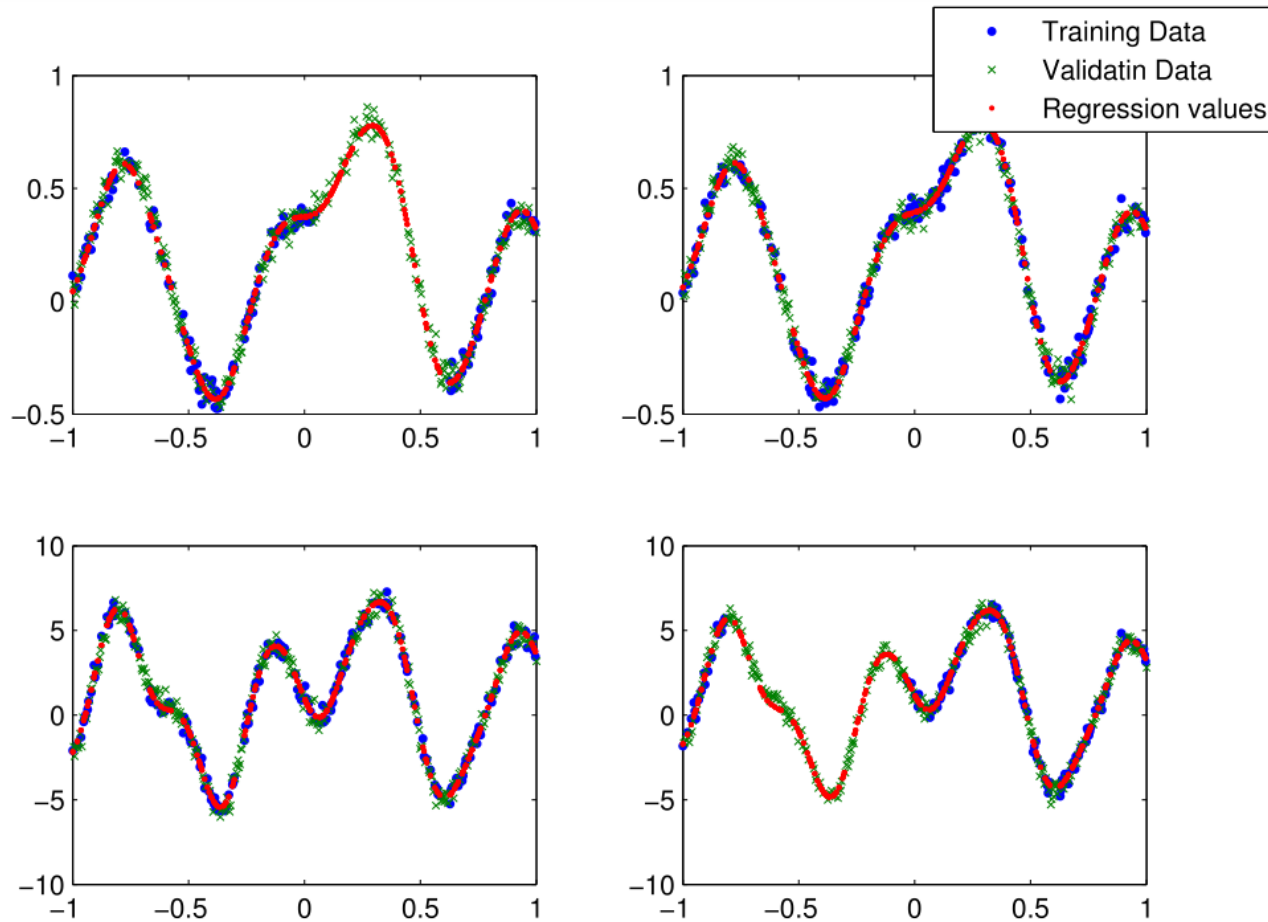
- داده‌ی تست از مقاله‌ی [1]، پیاده‌سازی الگوریتم‌ها در MATLAB
- آموزش با GP‌های مستقل از هم



[1] Alvarez, M. and Lawrence, N. Sparse convolved gaussian processes for multi-output regression. 2008

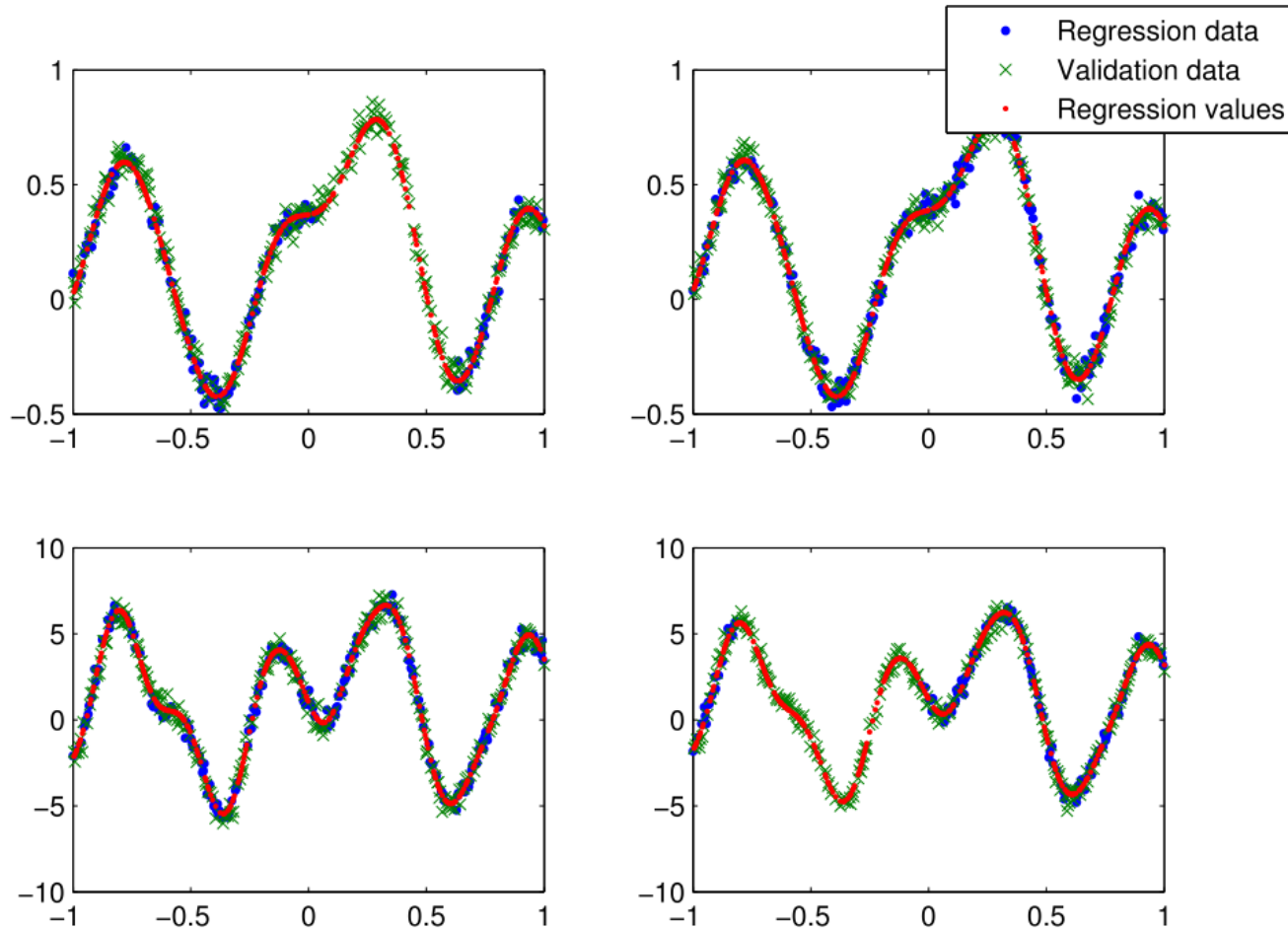
نتایج (۲)

• آموزش با مدل Spike and Slab



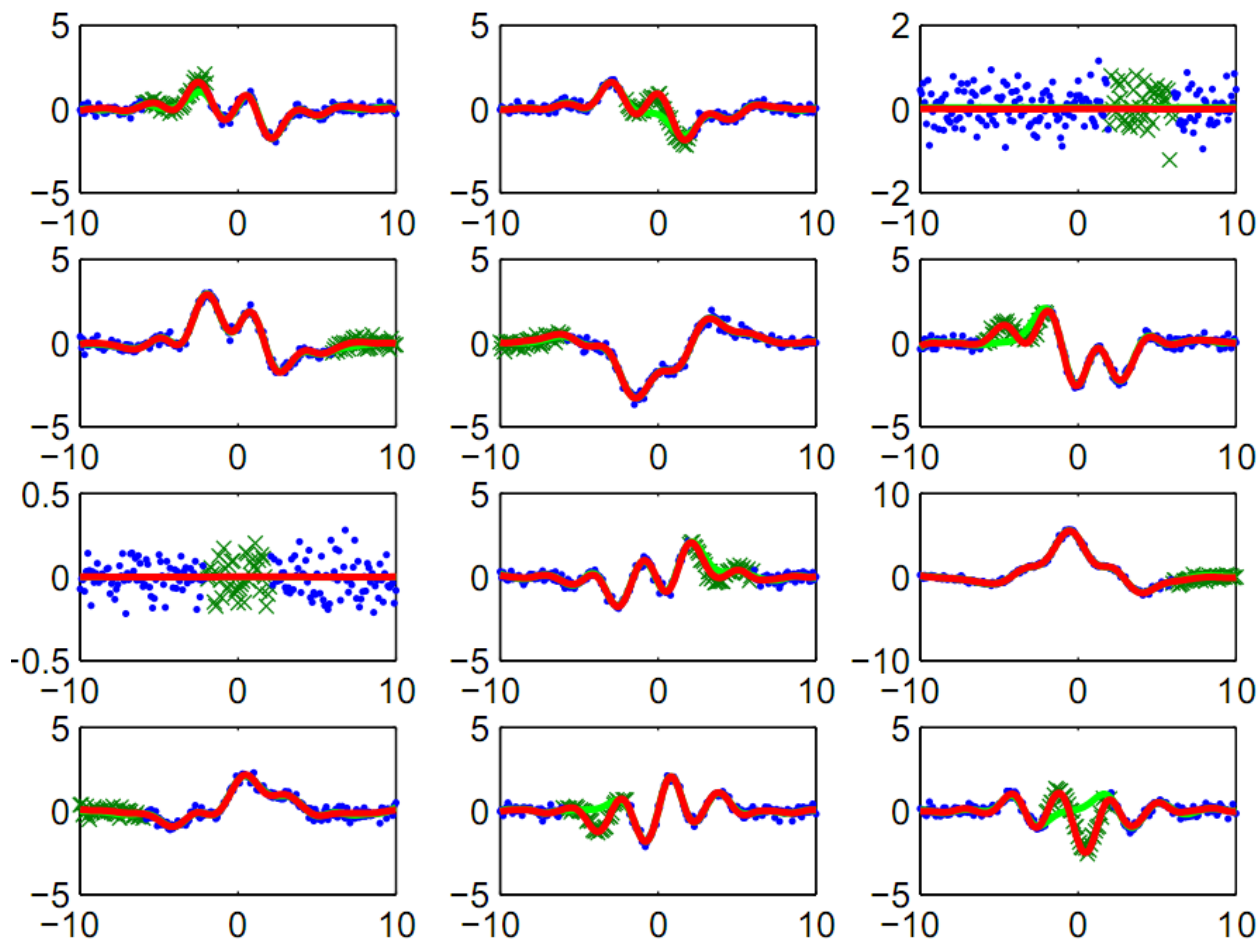
نتایج (۳)

• آموزش با مدل کانولوشنی



نتایج (۴)

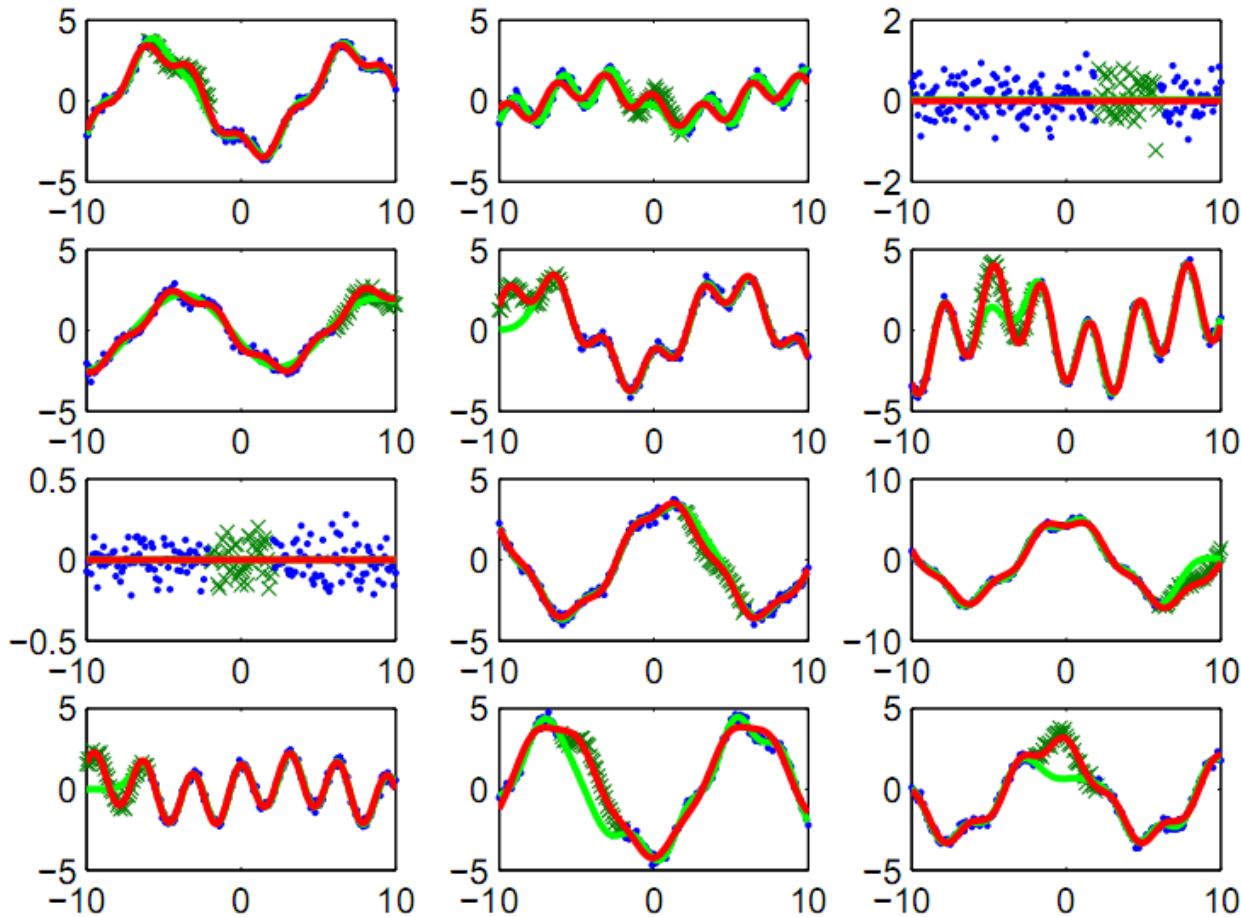
- آزمایشی دیگر با داده های [1]، نقاط آبی : داده های آموزشی، نقاط سبز: داده های validation. خط قرمز: مدل Spike and Slab، خط سبز: GP های مستقل



[1] Titsias, M.K. and L, M. Spike and slab variational inference for multi-task and multiple kernel learning, Advances in Neural Information Processing Systems(NIPS 2011), 2011.

نتایج (۵)

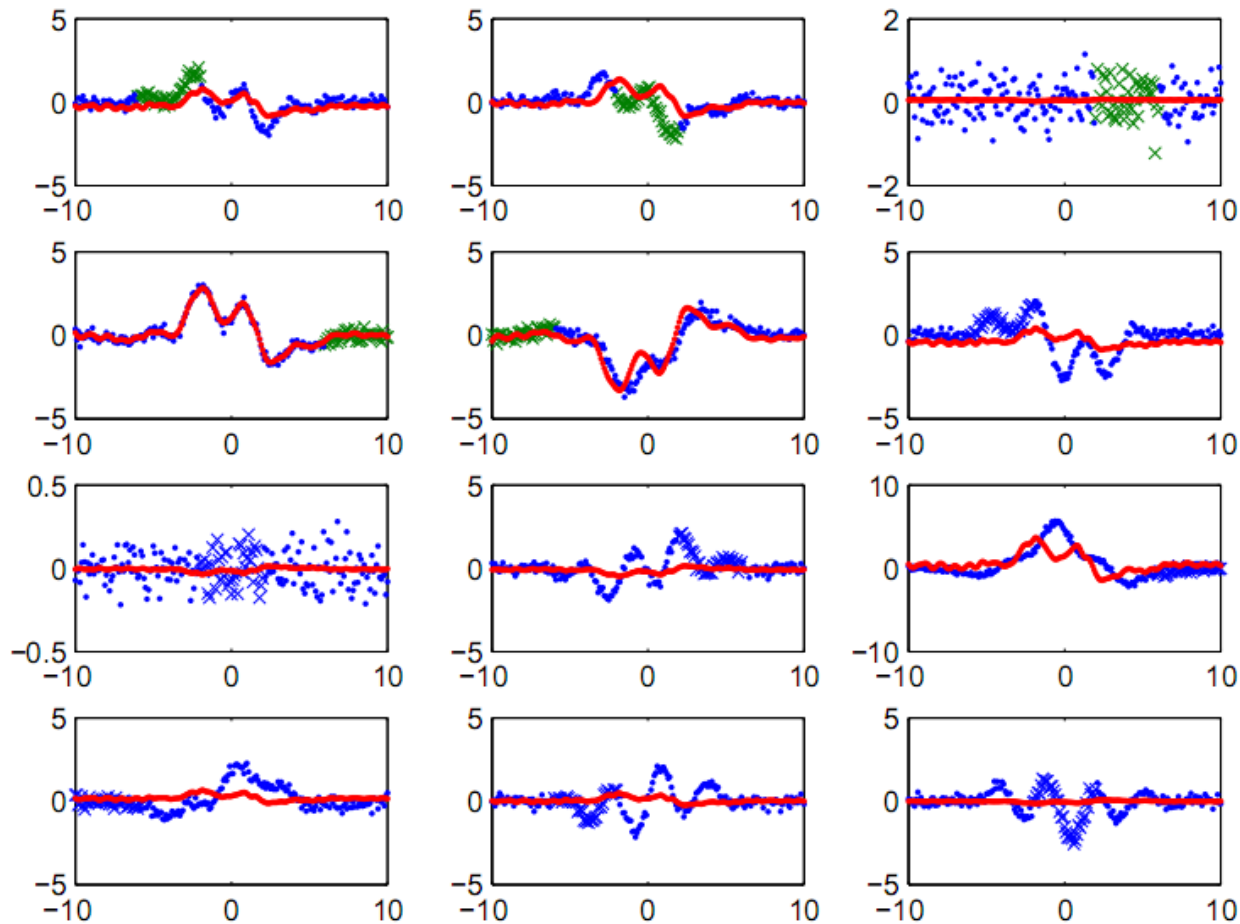
- آزمایشی دیگر با داده های [1]، نقاط آبی : داده های آموزشی، نقاط سبز: داده های validation. خط قرمز: مدل Spike and Slab، خط سبز: GP های مستقل



[1] Titsias, M.K. and L, M. Spike and slab variational inference for multi-task and multiple kernel learning, Advances in Neural Information Processing Systems(NIPS 2011), 2011.

نتایج (۶)

- آزمایشی دیگر با داده های [1]، نقاط آبی : داده های آموزشی، نقاط سبز: دادهای validation، خط قرمز: مدل کانولوشنی



[1] Titsias, M.K. and L, M. Spike and slab variational inference for multi-task and multiple kernel learning, Advances in Neural Information Processing Systems(NIPS 2011), 2011.

نتیجه گیری



- کارهایی که انجام شد:
 - بررسی مدل های یادگیری بیزوی معمولی
 - تعمیم یادگیری بیزوی معمولی برای مدل سازی متغیرهای همبسته با استفاده از دو روش اخیر
- کارهای آینده:
 - توانایی الگوریتم ها در مدل سازی هر نوع داده ها (ظرفیت مدل ها)
 - مدل سازی همبستگی های نیازمند بررسی ریاضی بیشتری است.
 - چه موقع باید همبستگی داشته باشیم و چه موقع نداشته باشیم؟
 - هر مدل به چه صورت از داده های مختلف برای بازسازی داده های از دسته رفته استفاده می کند؟
 - انواع دیگر همبستگی ها: مثل تاخیر
 - یادگیری بین ابعادی برای سایر مدل های یادگیری بیزوی
 - استفاده از مدل های معرفی شده در آزمایشات عملی تر برای واضح تر شدن قدرت آنها در استفاده های واقعی

تقدیر و تشکر:

- دکتر حمید شیخ زاده نجار
- سرکار خانم مهندس نجمه بطحایی
- سایر دوستان در آزمایشگاه MSPRL و دوستان در دانشکده ی برق

با تشکر از توجه شما!