

عیب‌یابی یک الگوریتم یادگیری ماشین

سید ناصر رضوی www.snrazavi.ir

۱۳۹۶

□ فرض کنید به منظور پیش‌بینی قیمت خانه‌ها الگوریتم رگرسیون خطی تنظیم شده را پیاده‌سازی نموده‌اید:

$$J(\theta) = \frac{1}{2} \left[\sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2 + \lambda \sum_{j=1}^n \theta_j^2 \right]$$

□ اما زمانی که فرضیه به دست آمده را بر روی یک مجموعه جدید از خانه‌ها آزمایش می‌کنید، متوجه می‌شوید که این فرضیه در پیش‌بینی خود شامل خطاهای بزرگ و غیرقابل قبولی است.

□ س. چگونه می‌توان این مشکل را برطرف نمود؟

□ راه‌حل‌های احتمالی.

□ از نمونه‌های آموزشی بیشتری استفاده کنید.

□ از تعداد ویژگی‌های کمتری استفاده کنید.

□ از تعداد ویژگی‌های بیشتری استفاده کنید.

□ سعی کنید ویژگی‌های چند جمله‌ای را به مجموعه‌ی ویژگی‌ها اضافه کنید. $(x_1^2, x_2^2, x_1x_2, \dots)$

□ ضریب تنظیم λ را کاهش دهید.

□ ضریب تنظیم λ را افزایش دهید.

عیب‌یابی یک سیستم یادگیری ماشین

۴

□ عیب‌یابی.

آزمایشی که با اجرای آن می‌توانید بفهمید کدام جنبه‌های یک الگوریتم یادگیری به درستی عمل نمی‌کنند و چگونه می‌توان عملکرد الگوریتم یادگیری را به بهترین شکل ممکن بهبود بخشید.

□ اگرچه پیاده‌سازی روش‌های عیب‌یابی ممکن است زمان‌بر باشند، اما استفاده از این روش‌ها در نهایت باعث صرفه‌جویی قابل ملاحظه‌ای در وقت شما خواهد شد.

ارزیابی فرضیه

۵

ارزیابی فرضیه

۶

□ عدم قابلیت تعمیم.

پاسخ نامناسب برای نمونه‌های جدیدی که قبلاً آموزش داده نشده‌اند.

□ ویژگی‌ها.

□ x_1 : اندازه خانه

□ x_2 : تعداد اتاق خواب‌ها

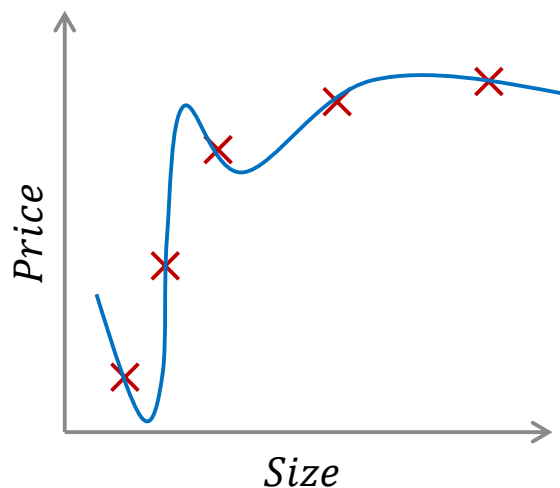
□ x_3 : تعداد طبقات

□ x_4 : قدمت

□ x_5 : اندازه آشپزخانه

□ ...

□ x_{100} : میانگین درآمد همسایه‌ها

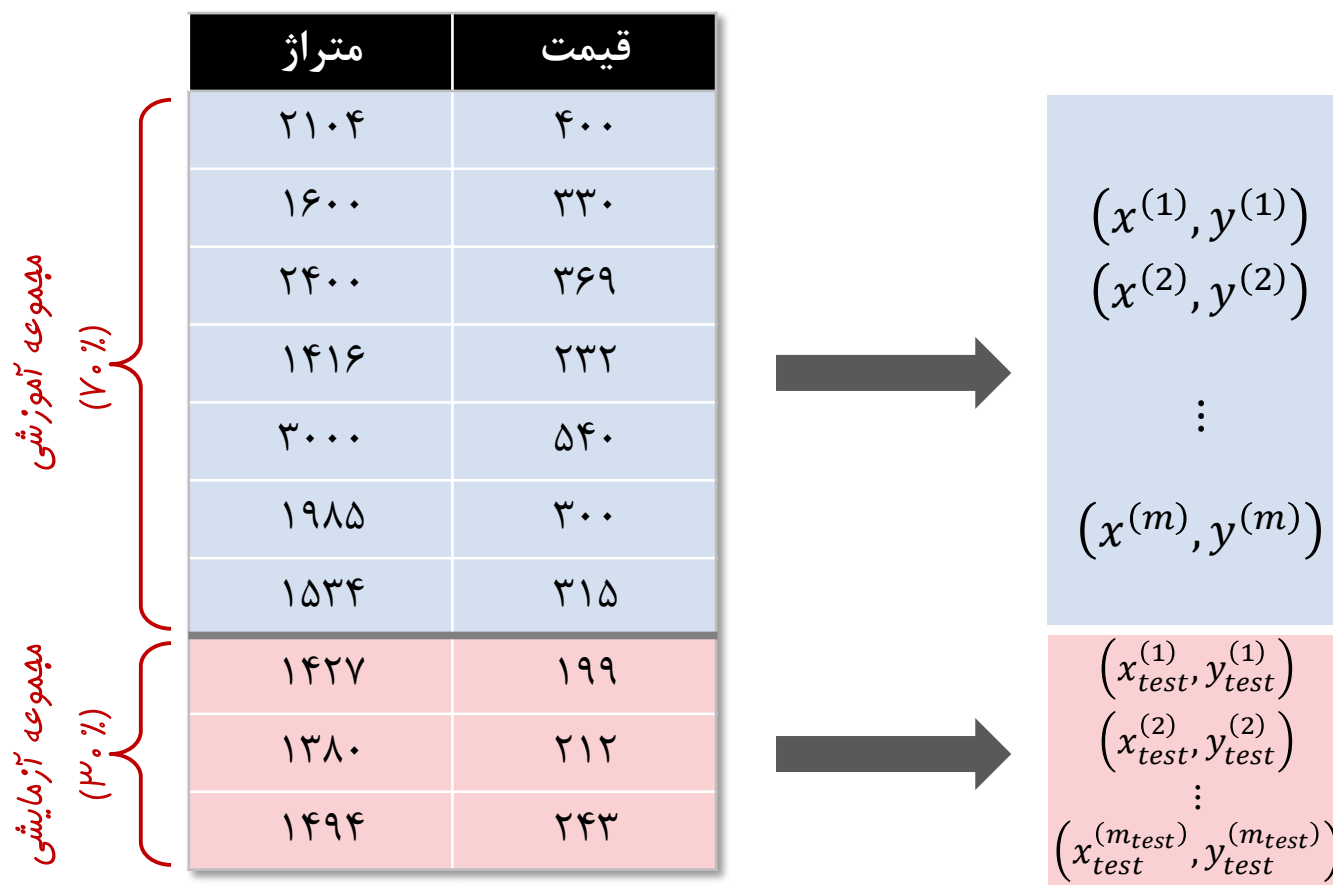


$$h_{\theta}(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$$

ارزیابی فرضیه

۷

□ مجموعه داده‌ها.



آموزش و آزمایش برای رگرسیون خطی

۸

□ آموزش.

یادگیری پارامترهای θ با استفاده از مجموعه آموزشی با کمینه کردن تابع هزینه $J(\theta)$.

$$J(\theta) = \frac{1}{2} \left[\sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2 + \lambda \sum_{j=1}^n \theta_j^2 \right]$$

□ آزمایش.

محاسبه خطا برای مجموعه آزمایشی.

$$J_{test}(\theta) = \frac{1}{2} \sum_{i=1}^{m_{test}} (h_{\theta}(x^{(i)}) - y^{(i)})^2$$

آموزش و آزمایش برای رگرسیون لجستیک

۹

□ آموزش.

یادگیری پارامترهای θ با استفاده از مجموعه آموزشی با کمینه کردن تابع هزینه $J(\theta)$.

$$J(\theta) = - \sum_{i=1}^m y^{(i)} \log h_{\theta}(x^{(i)}) + (1 - y^{(i)}) \log (1 - h_{\theta}(x^{(i)}))$$

□ آزمایش.

محاسبه خطا برای مجموعه آزمایشی.

$$J_{test}(\theta) = - \sum_{i=1}^{m_{test}} y_{test}^{(i)} \log h_{\theta}(x_{test}^{(i)}) + (1 - y_{test}^{(i)}) \log (1 - h_{\theta}(x_{test}^{(i)}))$$

آموزش و آزمایش برای رگرسیون لجستیک

۱۰

□ آموزش.

یادگیری پارامترهای θ با استفاده از مجموعه آموزشی با کمینه کردن تابع هزینه $J(\theta)$.

$$J(\theta) = - \sum_{i=1}^m y^{(i)} \log h_{\theta}(x^{(i)}) + (1 - y^{(i)}) \log (1 - h_{\theta}(x^{(i)}))$$

□ آزمایش. محاسبه **خطای کلاس‌بندی** برای مجموعه آزمایشی.

$$J_{test}(\theta) = \sum_{i=1}^{m_{test}} err(h_{\theta}(x^{(i)}), y^{(i)})$$

$$err(h_{\theta}(x), y) = \begin{cases} 1 & h_{\theta}(x) < 0.5, y = 1 \\ 1 & h_{\theta}(x) \geq 0.5, y = 0 \\ 0 & otherwise \end{cases}$$

انتخاب مدل

۱۱

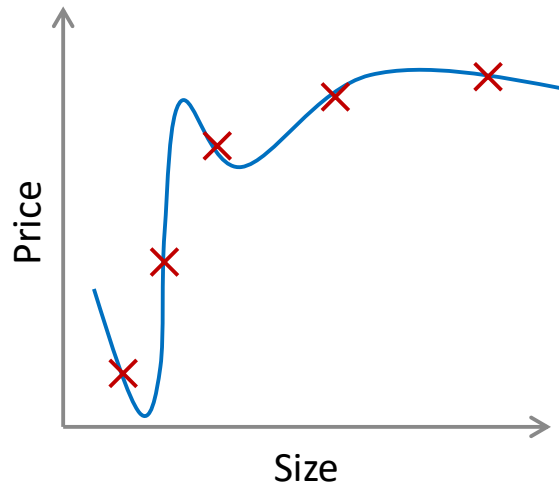
مجموعه آموزشی

مجموعه اعتبارسنجی

مجموعه آزمایشی

مثال بیش‌برازش

۱۲



$$h_{\theta}(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$$

□ پس از یاد گرفتن مقدار پارامترها از روی یک مجموعه آموزشی، خطای محاسبه شده بر روی مجموعه آموزشی معمولاً کمتر از مقدار خطای واقعی تعمیم است.

□ به عبارت دیگر، پایین بودن خطای آموزشی لزوماً به معنای مناسب بودن فرضیه نیست.

□ س. چگونه می‌توان خطای تعمیم را تخمین زد؟

انتخاب مدل

۱۳

□ س. کدام یک از مدل‌های زیر برای یک مجموعه از داده‌های مفروض بهتر است؟

$$h_{\theta}(x) = \theta_0 + \theta_1 x$$

$$h_{\theta}(x) = \theta_0 + \theta_1 x + \theta_2 x^2$$

$$h_{\theta}(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3$$

⋮

$$h_{\theta}(x) = \theta_0 + \theta_1 x + \cdots + \theta_{10} x^{10}$$

$$\min_{\theta} J(\theta) \rightarrow \theta^{(1)} \rightarrow J_{test}(\theta^{(1)})$$

$$\min_{\theta} J(\theta) \rightarrow \theta^{(2)} \rightarrow J_{test}(\theta^{(2)})$$

$$\min_{\theta} J(\theta) \rightarrow \theta^{(3)} \rightarrow J_{test}(\theta^{(3)})$$

⋮

$$\min_{\theta} J(\theta) \rightarrow \theta^{(10)} \rightarrow J_{test}(\theta^{(10)})$$

□ انتخاب مدل. مدلی را انتخاب کنید که کمترین خطای آزمایشی را دارد. فرض کنید چند جمله‌ای درجه ۵ دارای کمترین خطای آزمایشی باشد.

~~$J_{test}(\theta^{(5)})$~~

□ تخمین قابلیت تعمیم. قابلیت تعمیم مدل انتخاب شده چقدر است؟

ارزیابی فرضیه

۱۴

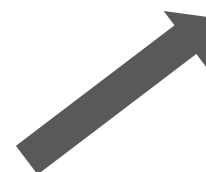
□ مجموعه داده‌ها.

مترائز	قیمت
۲۱۰۴	۴۰۰
۱۶۰۰	۳۳۰
۲۴۰۰	۳۶۹
۱۴۱۶	۲۳۲
۳۰۰۰	۵۴۰
۱۹۸۵	۳۰۰
۱۵۳۴	۳۱۵
۱۴۲۷	۱۹۹
۱۳۸۰	۲۱۲
۱۴۹۴	۲۴۳

مجموعه آموزشی (۶۰٪)

مجموعه اعتبار سنجی (۲۰٪)

مجموعه آزمایشی (۲۰٪)



$$\begin{pmatrix} x_{cv}^{(1)}, y_{cv}^{(1)} \\ x_{cv}^{(2)}, y_{cv}^{(2)} \\ \vdots \\ x_{cv}^{(m_{cv})}, y_{cv}^{(m_{cv})} \end{pmatrix}$$

مجموعه آموزشی / اعتبارسنجی / آزمایشی

۱۵

$$J_{train}(\theta) = \frac{1}{2} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2$$

□ مجموعه آموزشی.

یادگیری مقدار پارامترهای مدل

$$J_{cv}(\theta) = \frac{1}{2} \sum_{i=1}^{m_{cv}} (h_{\theta}(x_{cv}^{(i)}) - y_{cv}^{(i)})^2$$

□ مجموعه اعتبارسنجی.

انتخاب یک مدل [تعیین ابرپارامترها]

$$J_{test}(\theta) = \frac{1}{2} \sum_{i=1}^{m_{test}} (h_{\theta}(x_{test}^{(i)}) - y_{test}^{(i)})^2$$

□ مجموعه آزمایشی.

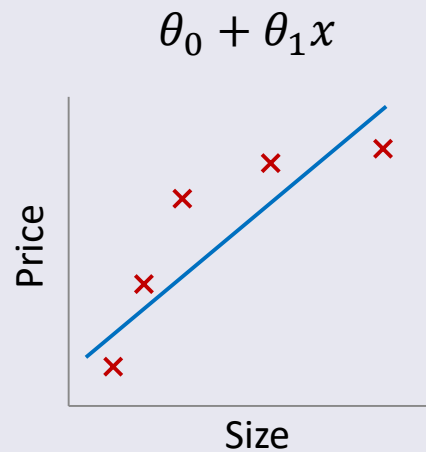
تخمین خطای تعمیم مدل انتخاب شده

تشخیص بایاس و واریانس بالا

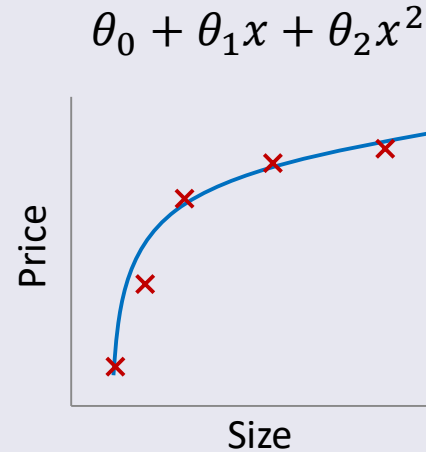
۱۶

یادآوری: بایاس و واریانس

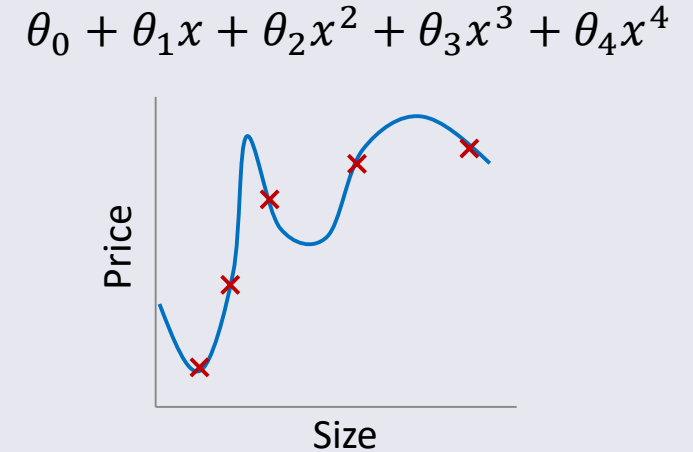
۱۷



کم برآزش (بایاس بالا)



مدل صحیح



بیش برآزش (واریانس بالا)

□ بیش برآزش. اگر تعداد ویژگی‌ها بسیار زیاد باشد، فرضیه یاد گرفته شده ممکن است داده‌های آموزشی را خیلی خوب یاد بگیرد، اما این امکان نیز وجود دارد که این فرضیه در پیش‌بینی داده‌های جدید شکست بخورد. [عدم قابلیت تعمیم]

بایاس و واریانس

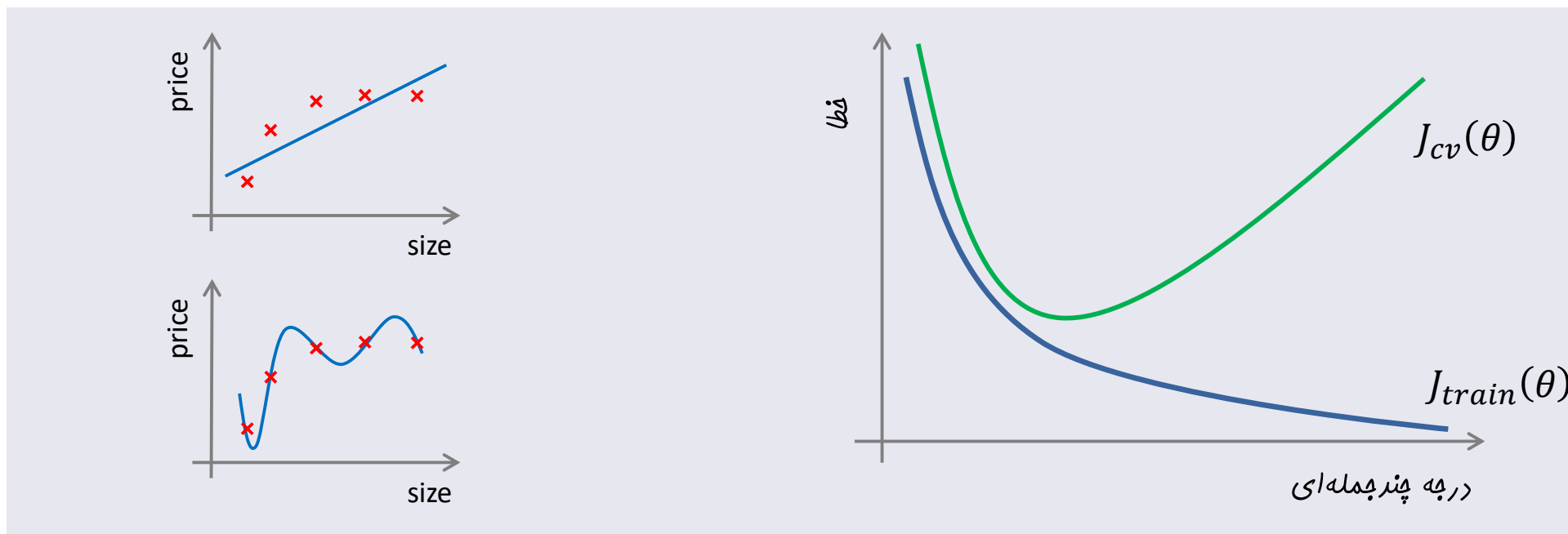
۱۸

$$J_{train}(\theta) = \frac{1}{2} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2$$

خطای آموزشی

$$J_{cv}(\theta) = \frac{1}{2} \sum_{i=1}^{m_{cv}} (h_{\theta}(x_{cv}^{(i)}) - y_{cv}^{(i)})^2$$

خطای اعتبارسنجی

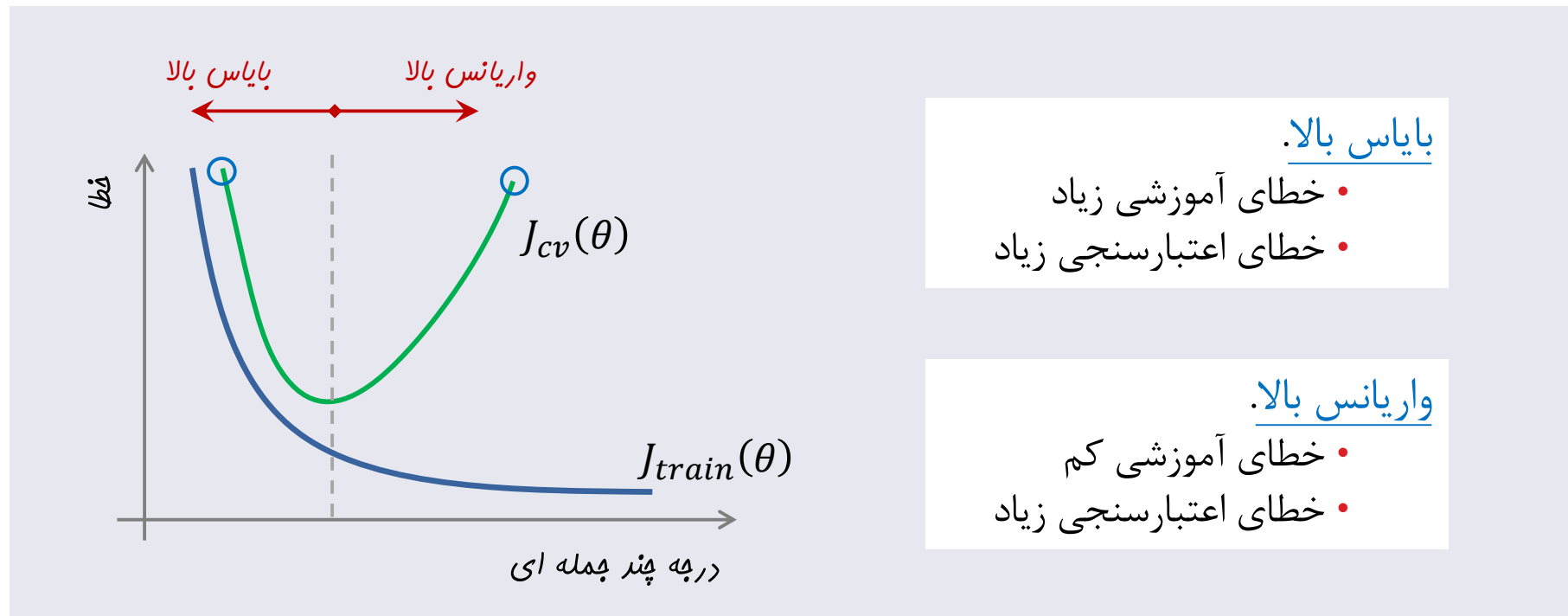


تشخیص بایاس در برابر واریانس

۱۹

□ فرض کنید کیفیت عملکرد الگوریتم شما کمتر از آن چیزی است که انتظار دارید! [خطای اعتبارسنجی زیاد است]

□ س. چگونه می‌توان تشخیص داد این مسئله ناشی از بایاس است یا واریانس؟



تنظیم و مشکل بایاس/واریانس

۲۰

□ تنظیم. استفاده از تنظیم می‌تواند در برخورد با مسئله بیش‌برازش مؤثر باشد.

$$J(\theta) = \frac{1}{2} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2 + \frac{\lambda}{2} \sum_{j=1}^n \theta_j^2$$

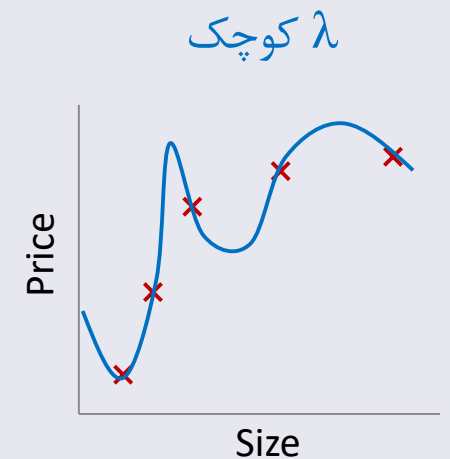
□ اما تنظیم چگونه بر روی بایاس و واریانس تأثیر می‌گذارد؟

انتخاب ضریب تنظیم λ

۲۲

$$h_{\theta}(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$$

$$J(\theta) = \frac{1}{2} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2 + \frac{\lambda}{2} \sum_{j=1}^n \theta_j^2$$



انتخاب ضریب تنظیم λ

۲۳

$$h_{\theta}(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$$

$$J(\theta) = \frac{1}{2} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2 + \frac{\lambda}{2} \sum_{j=1}^n \theta_j^2$$

$$J_{train}(\theta) = \frac{1}{2} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2$$

$$J_{cv}(\theta) = \frac{1}{2} \sum_{i=1}^{m_{cv}} (h_{\theta}(x_{cv}^{(i)}) - y_{cv}^{(i)})^2$$

$$J_{test}(\theta) = \frac{1}{2} \sum_{i=1}^{m_{test}} (h_{\theta}(x_{test}^{(i)}) - y_{test}^{(i)})^2$$

انتخاب ضریب تنظیم λ

۲۴

$$h_{\theta}(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$$

$$J(\theta) = \frac{1}{2} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2 + \frac{\lambda}{2} \sum_{j=1}^n \theta_j^2$$

- | | |
|----------------------|---|
| 1. $\lambda = 0.00$ | $\min_{\theta} J(\theta) \rightarrow \theta^{(1)} \rightarrow J_{cv}(\theta^{(1)})$ |
| 2. $\lambda = 0.01$ | $\min_{\theta} J(\theta) \rightarrow \theta^{(2)} \rightarrow J_{cv}(\theta^{(2)})$ |
| 3. $\lambda = 0.02$ | $\min_{\theta} J(\theta) \rightarrow \theta^{(3)} \rightarrow J_{cv}(\theta^{(3)})$ |
| 4. $\lambda = 0.04$ | $\min_{\theta} J(\theta) \rightarrow \theta^{(4)} \rightarrow J_{cv}(\theta^{(4)})$ |
| $\vdots$ | |
| 12. $\lambda = 10.0$ | $\min_{\theta} J(\theta) \rightarrow \theta^{(12)} \rightarrow J_{cv}(\theta^{(12)})$ |

انتخاب مدلی که کمترین
خطای اعتبارسنجی را دارد

انتخاب ضریب تنظیم λ

۲۵

$$h_{\theta}(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$$

$$J(\theta) = \frac{1}{2} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2 + \frac{\lambda}{2} \sum_{j=1}^n \theta_j^2$$

- | | |
|----------------------|---|
| 1. $\lambda = 0.00$ | $\min_{\theta} J(\theta) \rightarrow \theta^{(1)} \rightarrow J_{cv}(\theta^{(1)})$ |
| 2. $\lambda = 0.01$ | $\min_{\theta} J(\theta) \rightarrow \theta^{(2)} \rightarrow J_{cv}(\theta^{(2)})$ |
| 3. $\lambda = 0.02$ | $\min_{\theta} J(\theta) \rightarrow \theta^{(3)} \rightarrow J_{cv}(\theta^{(3)})$ |
| 4. $\lambda = 0.04$ | $\min_{\theta} J(\theta) \rightarrow \theta^{(4)} \rightarrow J_{cv}(\theta^{(4)})$ |
| $\vdots$ | |
| 12. $\lambda = 10.0$ | $\min_{\theta} J(\theta) \rightarrow \theta^{(12)} \rightarrow J_{cv}(\theta^{(12)})$ |

انتخاب مدلی که کمترین
خطای اعتبارسنجی را دارد

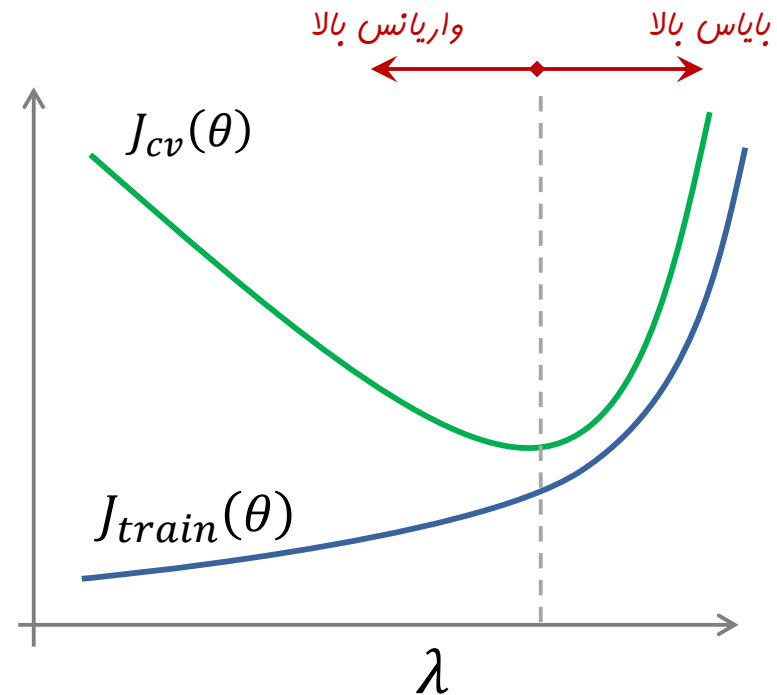
انتخاب ضریب تنظیم λ

۲۶

$$J(\theta) = \frac{1}{2} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2 + \frac{\lambda}{2} \sum_{j=1}^n \theta_j^2$$

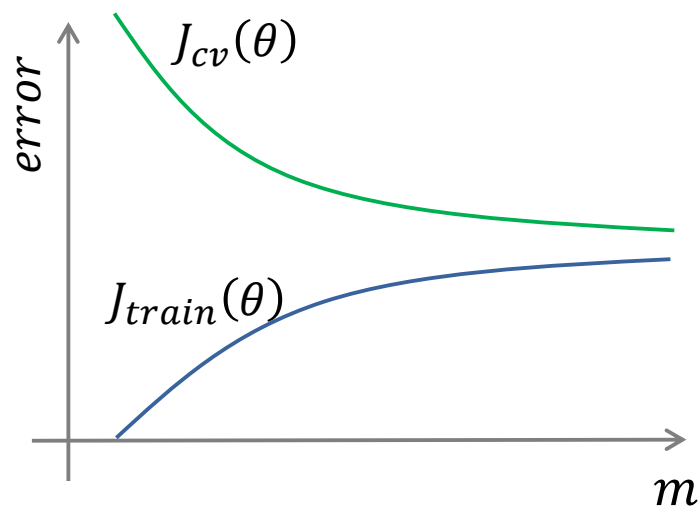
$$J_{train}(\theta) = \frac{1}{2} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2$$

$$J_{cv}(\theta) = \frac{1}{2} \sum_{i=1}^{m_{cv}} (h_{\theta}(x_{cv}^{(i)}) - y_{cv}^{(i)})^2$$

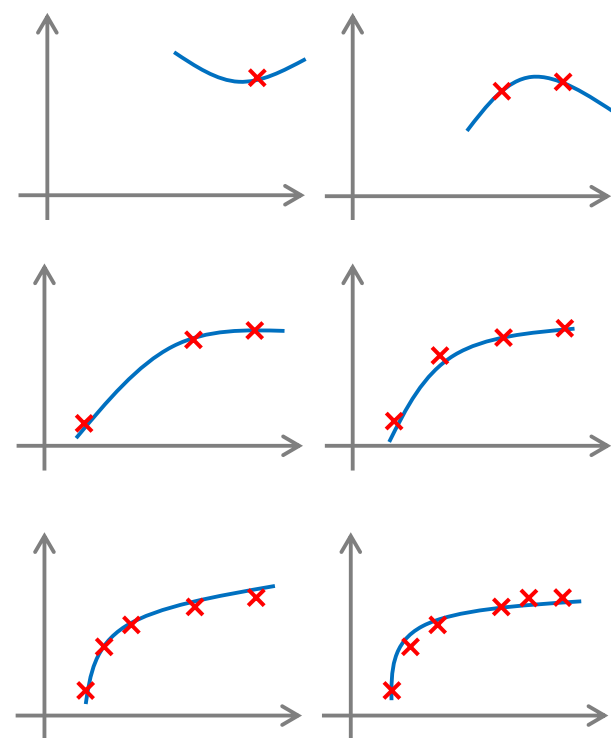


$$J_{train}(\theta) = \frac{1}{2} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2$$

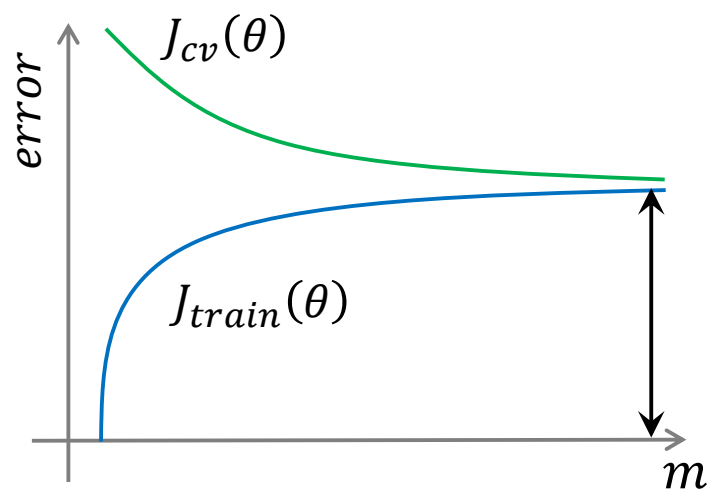
$$J_{cv}(\theta) = \frac{1}{2} \sum_{i=1}^{m_{cv}} (h_{\theta}(x_{cv}^{(i)}) - y_{cv}^{(i)})^2$$



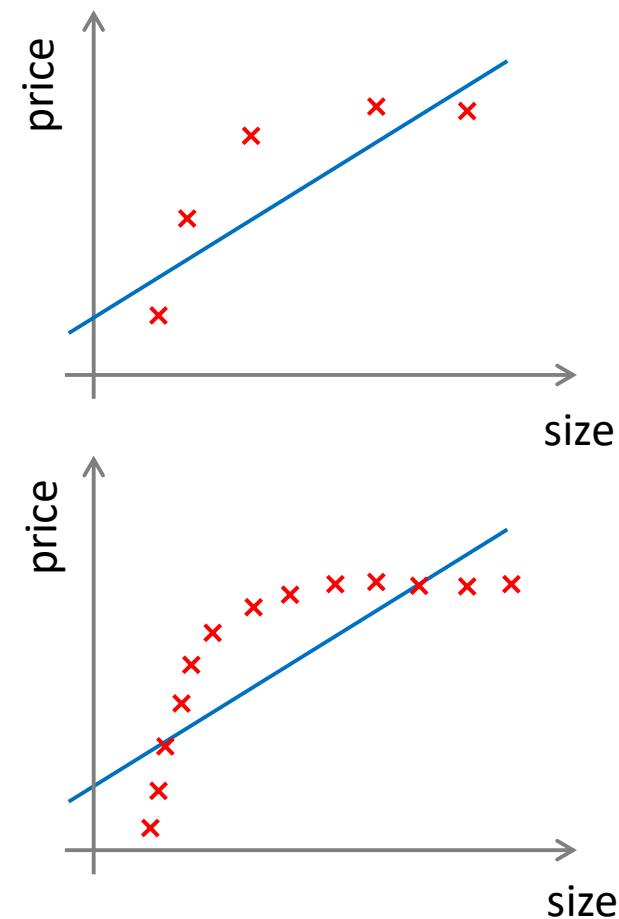
$$h_{\theta}(x) = \theta_0 + \theta_1 x + \theta_2 x^2$$



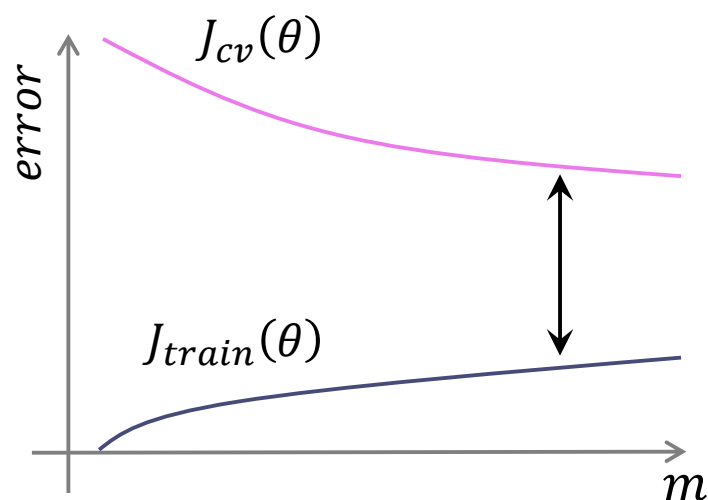
$$h_{\theta}(x) = \theta_0 + \theta_1 x$$



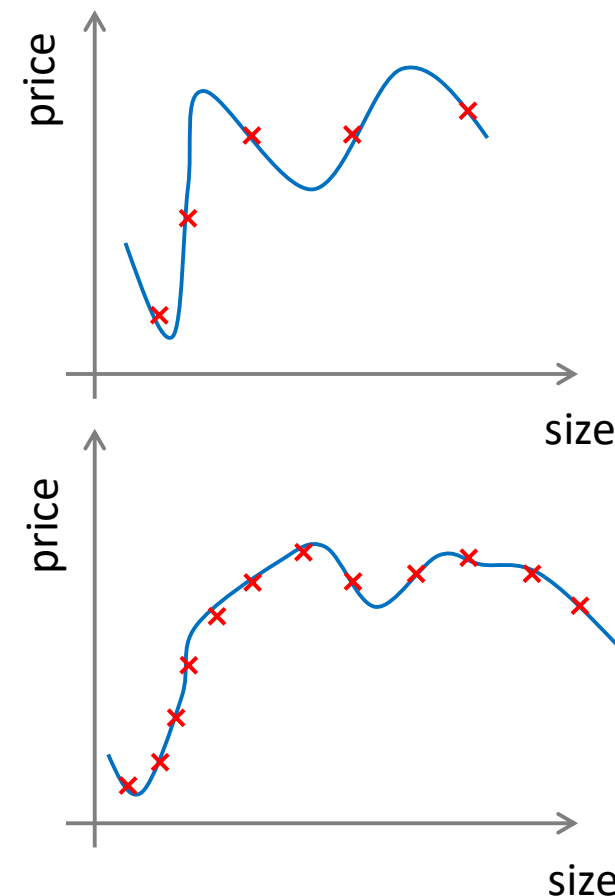
اگر یک الگوریتم یادگیری از **بایاس بالا** رنج ببرد، افزایش تعداد نمونه‌های آموزشی کمک چندانی به آن نخواهد کرد.



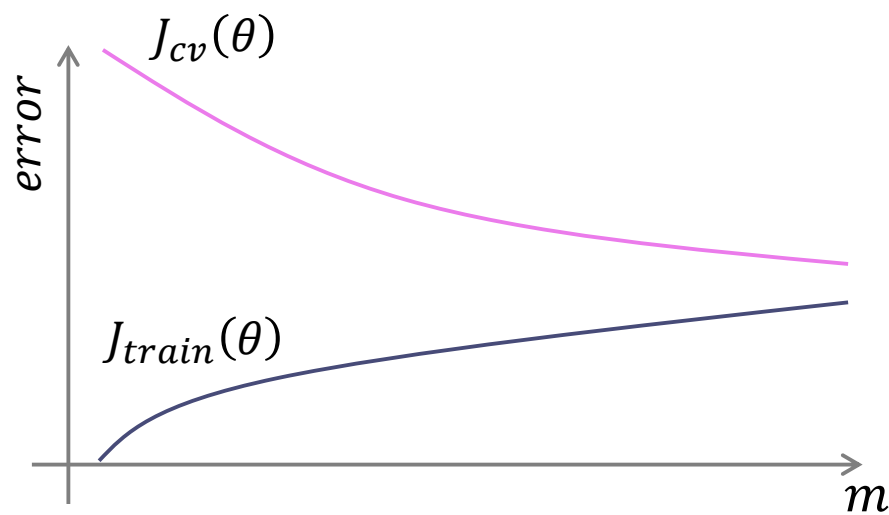
$$h_{\theta}(x) = \theta_0 + \theta_1 x + \dots + \theta_{100} x^{100}$$



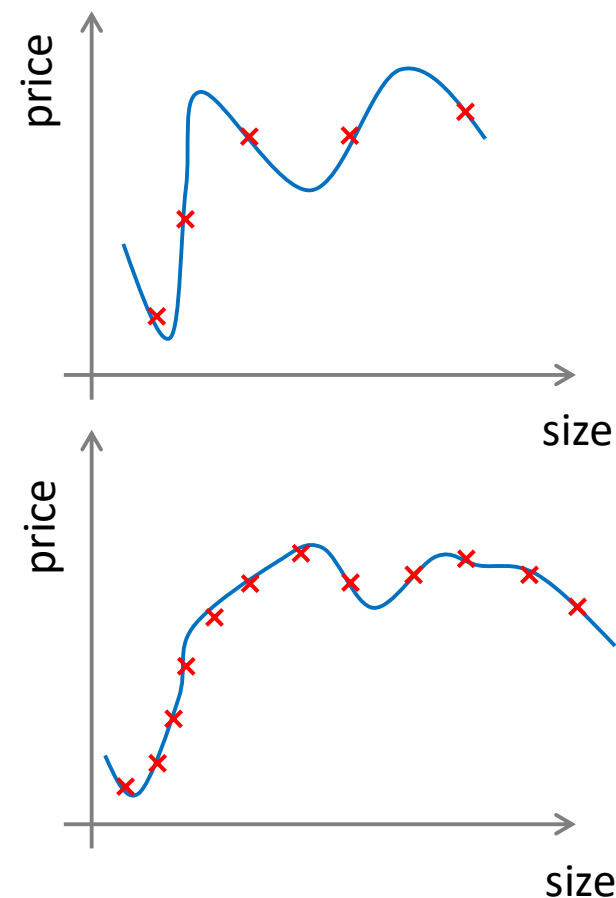
اگر یک الگوریتم یادگیری از واریانس بالا رنج
ببرد، افزایش تعداد نمونه‌های آموزشی احتمالاً
کمک کننده خواهد بود.



$$h_{\theta}(x) = \theta_0 + \theta_1 x + \dots + \theta_{100} x^{100}$$



اگر یک الگوریتم یادگیری از واریانس بالا رنج
ببرد، افزایش تعداد نمونه‌های آموزشی احتمالاً
کمک کننده خواهد بود.



□ فرض کنید به منظور پیش‌بینی قیمت خانه‌ها الگوریتم رگرسیون خطی تنظیم شده را پیاده‌سازی نموده‌اید:

$$J(\theta) = \frac{1}{2} \left[\sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2 + \lambda \sum_{j=1}^n \theta_j^2 \right]$$

□ اما زمانی که فرضیه به دست آمده را بر روی یک مجموعه جدید از خانه‌ها آزمایش می‌کنید، متوجه می‌شوید که این فرضیه در پیش‌بینی خود شامل خطاهای بزرگ و غیرقابل قبولی است.

□ س. چگونه می‌توان این مشکل را برطرف نمود؟

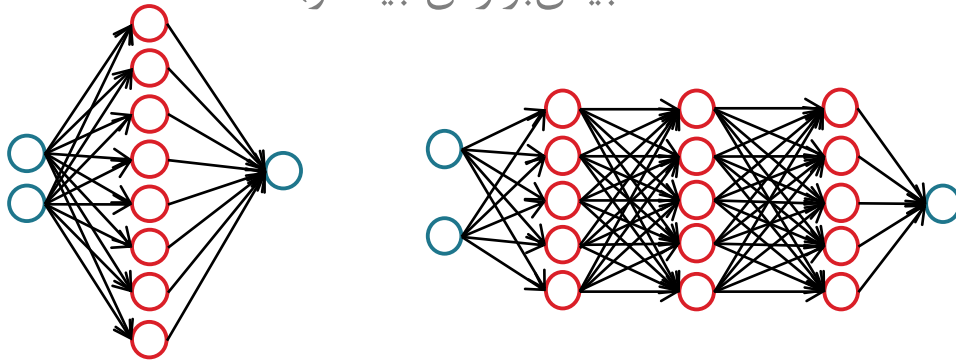
□ راه‌حل‌های احتمالی.

- | | | |
|---------------------------------|---|-----------------------|
| □ افزایش تعداد نمونه‌های آموزشی | ← | حل مسئله واریانس بالا |
| □ کاهش تعداد ویژگی‌ها | ← | حل مسئله واریانس بالا |
| □ افزایش تعداد ویژگی‌ها | ← | حل مسئله بایاس بالا |
| □ افزودن ویژگی‌های چندجمله‌ای | ← | حل مسئله بایاس بالا |
| □ کاهش ضریب تنظیم λ | ← | حل مسئله بایاس بالا |
| □ افزایش ضریب تنظیم λ | ← | حل مسئله واریانس بالا |

شبکه‌های عصبی و بیش‌برازش

۳۵

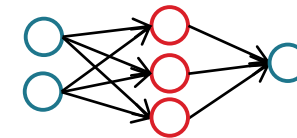
شبکه‌ی عصبی «بزرگ»
(پارامترهای بیشتر؛ احتمال
بیش‌برازش بیشتر)



هزینه محاسباتی بیشتر

استفاده از تنظیم برای برخورد با بیش‌برازش

شبکه‌ی عصبی «کوچک»
(پارامترهای کمتر؛ احتمال
کم‌برازش بیشتر)



هزینه محاسباتی کمتر