فصل هشتم : عیب یابی

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

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

- 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؟

فصل هشتم : عیب یابی

- راه‌حل‌های احتمالی.
  - از نمونه‌های آموزشی بیشتری استفاده کنید.
  - از تعداد ویژگی‌های کمتری استفاده کنید.
  - از تعداد ویژگی‌های بیشتری استفاده کنید.
  - سعی کنید ویژگی‌های چندجمله‌ای 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.

فصل هشتم : عیب یابی یک سیستم یادگیری ماشین

- 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.

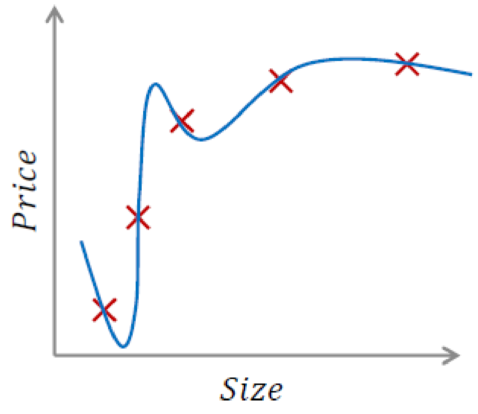
۵

ارزیابی فرضیه

فصل هشتم : ارزیابی فرضیه

- عدم قابلیت تعمیم.

پاسخ نامناسب برای نمونه‌های جدیدی که قبلاً آموزش داده نشده‌اند.



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

- 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

فصل هشتم : ارزیابی فرضیه

- مجموعه داده‌ها.

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

مجموعه آموزشی (۷۰٪)

مجموعه آزمایشی (۳۰٪)

$$\left(x^{(1)}, y^{(1)}\right)$$
$$\left(x^{(2)}, y^{(2)}\right)$$
$$\vdots$$
$$\left(x^{(m)}, y^{(m)}\right)$$

$$\left(x_{test}^{(1)}, y_{test}^{(1)}\right)$$
$$\left(x_{test}^{(2)}, y_{test}^{(2)}\right)$$
$$\vdots$$
$$\left(x_{test}^{(m_{test})}, y_{test}^{(m_{test})}\right)$$

فصل هشتم : آموزش و آزمایش برای رگرسیون خطی

- آموزش.

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

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

- آزمایش.

محاسبه خطا برای مجموعه آزمایشی.

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

فصل هشتم : آموزش و آزمایش برای رگرسیون لجستیک

- آموزش.

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

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

- آزمایش.

محاسبه خطا برای مجموعه آزمایشی.

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

فصل هشتم : آموزش و آزمایش برای رگرسیون لجستیک

- آموزش.

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

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

- آزمایش. محاسبه خطای کلاس‌بندی برای مجموعه آزمایشی.

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

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

انتخاب مدل

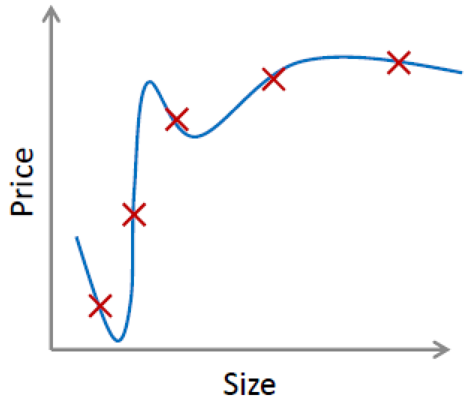
- مجموعه آموزشی
- مجموعه اعتبارسنجی
- مجموعه آزمایشی

فصل هشتم : مثال بیش برازش

- 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؟



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

فصل هشتم : انتخاب مدل

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

$$h_\theta(x) = \theta_0 + \theta_1 x \qquad \min_\theta J(\theta) \rightarrow \theta^{(1)} \rightarrow J_{test}(\theta^{(1)})$$

$$h_\theta(x) = \theta_0 + \theta_1 x + \theta_2 x^2 \qquad \min_\theta J(\theta) \rightarrow \theta^{(2)} \rightarrow J_{test}(\theta^{(2)})$$

$$h_\theta(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 \qquad \min_\theta J(\theta) \rightarrow \theta^{(3)} \rightarrow J_{test}(\theta^{(3)})$$

$$\vdots \qquad \vdots$$

$$h_\theta(x) = \theta_0 + \theta_1 x + \cdots + \theta_{10} x^{10} \qquad \min_\theta J(\theta) \rightarrow \theta^{(10)} \rightarrow J_{test}(\theta^{(10)})$$

- انتخاب مدل. مدلی را انتخاب کنید که کمترین خطای آزمایشی را دارد. فرض کنید چندجمله‌ای درجه ۵ دارای کمترین خطای آزمایشی باشد.
- تخمین قابلیت تعمیم. قابلیت تعمیم مدل انتخاب شده چقدر است؟

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

فصل هشتم : ارزیابی فرضیه

- مجموعه داده‌ها.

قیمت	متراژ	
۴۰۰	۲۱۰۴	مجموعه آموزشی (۶۰٪)
۳۳۰	۱۶۰۰	
۳۶۹	۲۴۰۰	
۲۳۲	۱۴۱۶	
۵۴۰	۳۰۰۰	
۳۰۰	۱۹۸۵	
۳۱۵	۱۵۳۴	مجموعه اعتبارسنجی (۲۰٪)
۱۹۹	۱۴۲۷	
۲۱۲	۱۳۸۰	مجموعه آزمایشی (۲۰٪)
۲۴۳	۱۴۹۴	

$$\left(x_{cv}^{(1)}, y_{cv}^{(1)}\right)$$
$$\left(x_{cv}^{(2)}, y_{cv}^{(2)}\right)$$
$$\vdots$$
$$\left(x_{cv}^{(m_{cv})}, y_{cv}^{(m_{cv})}\right)$$

فصل هشتم : مجموعه آموزشی /اعتبارسنجی /آزمایشی

- مجموعه آموزشی.

  یادگیری مقدار پارامترهای مدل

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

- مجموعه اعتبارسنجی.

  انتخاب یک مدل [تعیین ابرپارامترها]

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

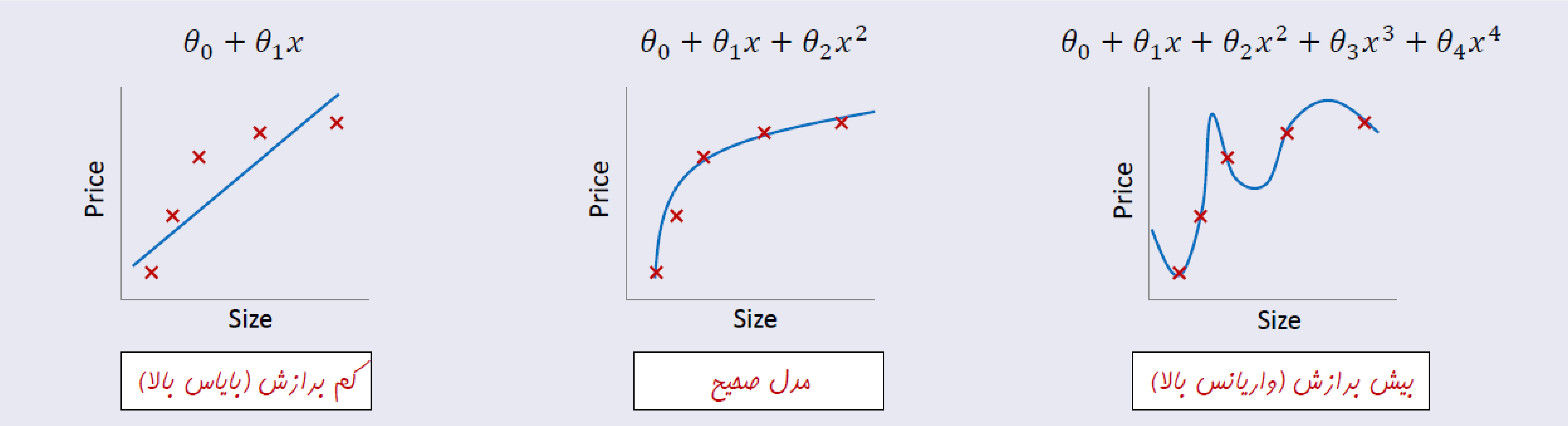
- مجموعه آزمایشی.

  تخمین خطای تعمیم مدل انتخاب شده

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

تشخیص بایاس و واریانس بالا

فصل هشتم : یادآوری(بایاس و واریانس)



- بیش‌برازش. 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]

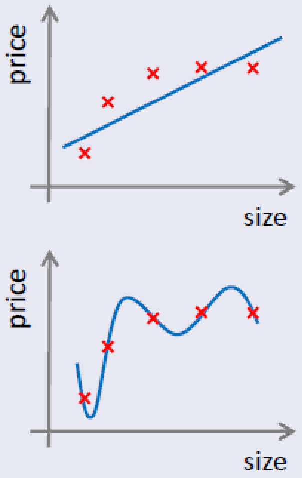
فصل هشتم : بایاس و واریانس

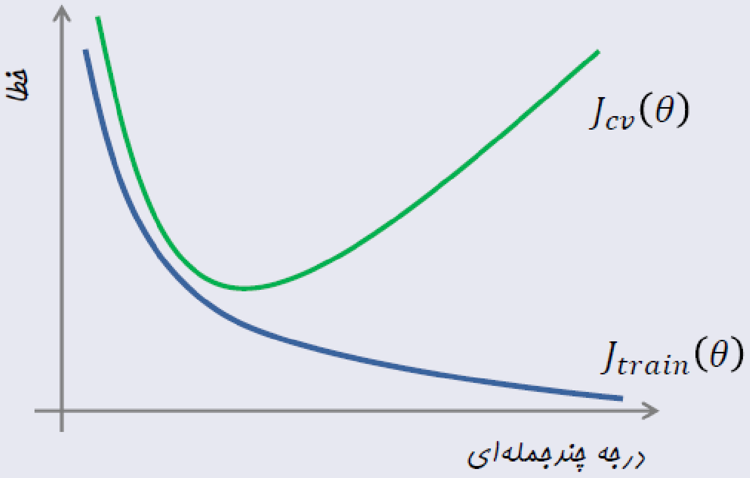
$$J_{cv}(\theta) = \frac{1}{2}\sum_{i=1}^{m_{cv}}\left(h_\theta\left(x_{cv}^{(i)}\right) - y_{cv}^{(i)}\right)^2$$

خطای اعتبارسنجی

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

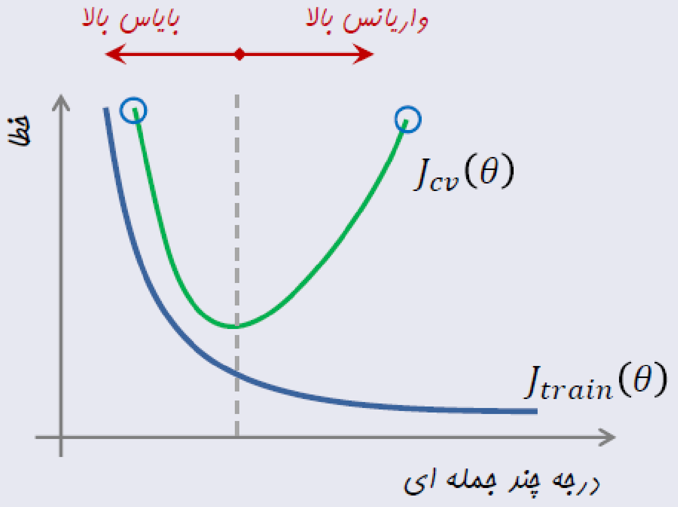
خطای آموزشی





فصل هشتم : تشخیص بایاس در برابر واریانس

- 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؟



بایاس بالا.
- خطای آموزشی زیاد
- خطای اعتبارسنجی زیاد

واریانس بالا.
- خطای آموزشی کم
- خطای اعتبارسنجی زیاد

تنظیم و مشکل بایاس/واریانس

فصل هشتم : تنظیم

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

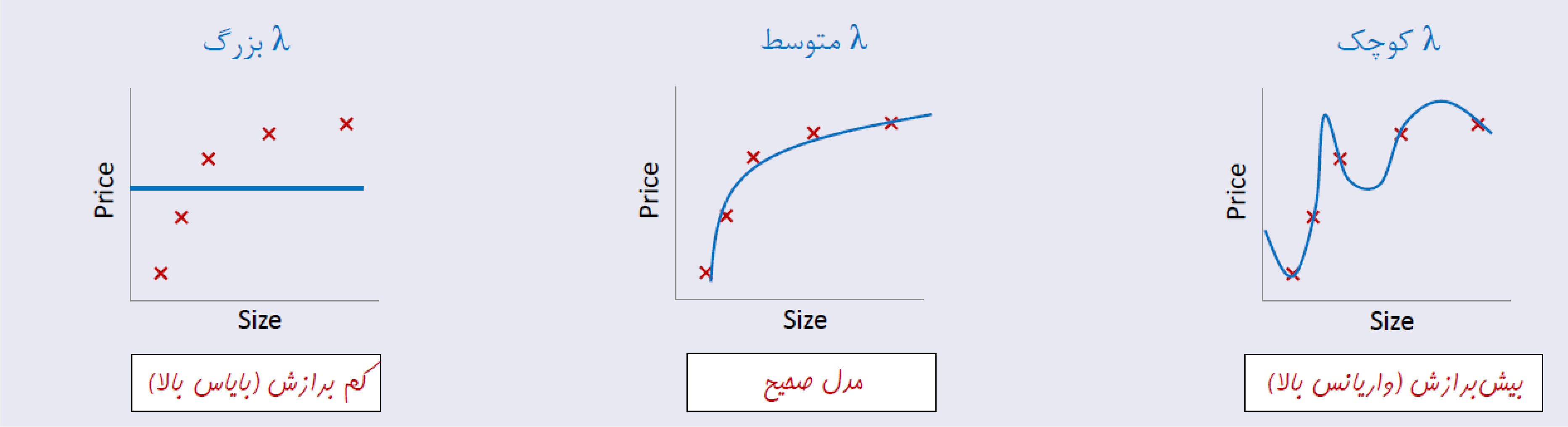
$$J(\theta) = \frac{1}{2}\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2 + \frac{\lambda}{2}\sum_{j=1}^{n}\theta_j^2$$

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

فصل هشتم : انتخاب ضریب تنظیم λ

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

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



فصل هشتم : انتخاب ضریب تنظیم ۸

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

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

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

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

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

فصل هشتم : انتخاب ضریب تنظیم λ

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

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

$$\left.\begin{array}{lll}
1.\ \lambda = 0.00 & \min\limits_\theta J(\theta) \to \theta^{(1)} & \to J_{cv}(\theta^{(1)}) \\
2.\ \lambda = 0.01 & \min\limits_\theta J(\theta) \to \theta^{(2)} & \to J_{cv}(\theta^{(2)}) \\
3.\ \lambda = 0.02 & \min\limits_\theta J(\theta) \to \theta^{(3)} & \to J_{cv}(\theta^{(3)}) \\
4.\ \lambda = 0.04 & \min\limits_\theta J(\theta) \to \theta^{(4)} & \to J_{cv}(\theta^{(4)}) \\
\vdots & & \\
12.\ \lambda = 10.0 & \min\limits_\theta J(\theta) \to \theta^{(12)} & \to J_{cv}(\theta^{(12)})
\end{array}\right\}$$

انتخاب مدلی که کمترین
خطای اعتبارسنجی را دارد

فصل هشتم : انتخاب ضریب تنظیم λ

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

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

1. $\lambda = 0.00$ $\quad \min_\theta J(\theta) \rightarrow \theta^{(1)} \rightarrow J_{cv}\left(\theta^{(1)}\right)$
2. $\lambda = 0.01$ $\quad \min_\theta J(\theta) \rightarrow \theta^{(2)} \rightarrow J_{cv}\left(\theta^{(2)}\right)$
3. $\lambda = 0.02$ $\quad \min_\theta J(\theta) \rightarrow \theta^{(3)} \rightarrow J_{cv}\left(\theta^{(3)}\right)$
4. $\lambda = 0.04$ $\quad \min_\theta J(\theta) \rightarrow \theta^{(4)} \rightarrow J_{cv}\left(\theta^{(4)}\right)$

$\vdots$

12. $\lambda = 10.0$ $\quad \min_\theta J(\theta) \rightarrow \theta^{(12)} \rightarrow J_{cv}\left(\theta^{(12)}\right)$

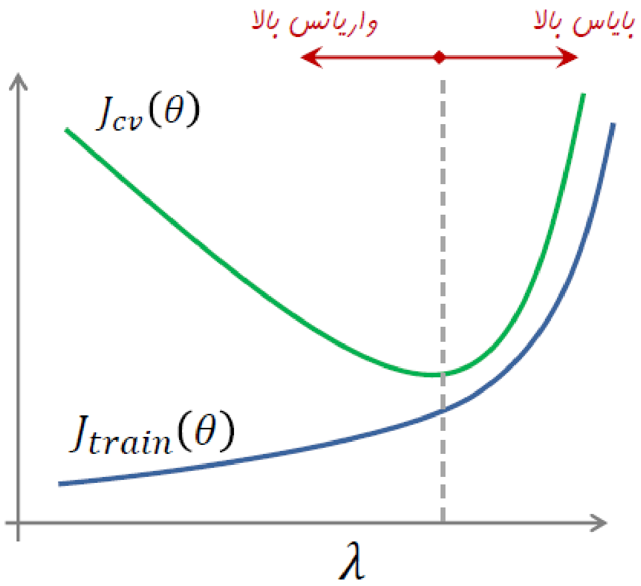
انتخاب مدلی که کمترین
خطای اعتبارسنجی را دارد

فصل هشتم : انتخاب ضریب تنظیم λ

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

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

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

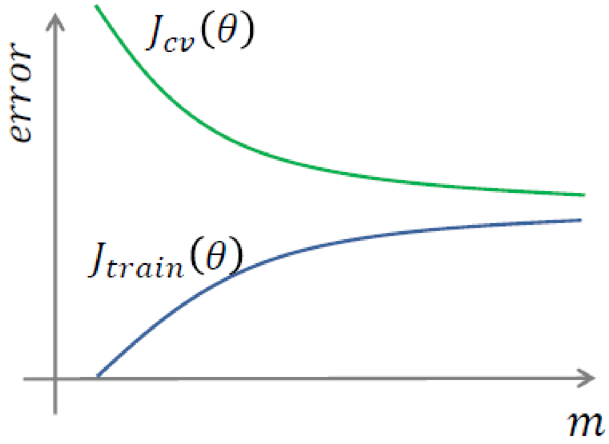


منحنی‌های یادگیری

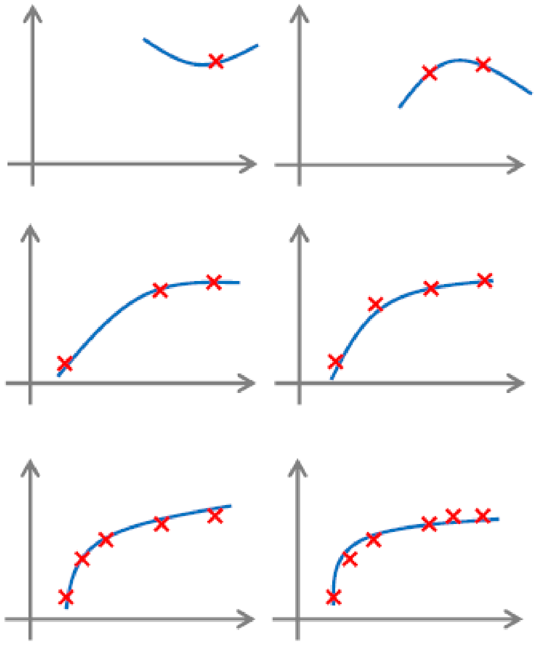
فصل هشتم : منحنی های یادگیری

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

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

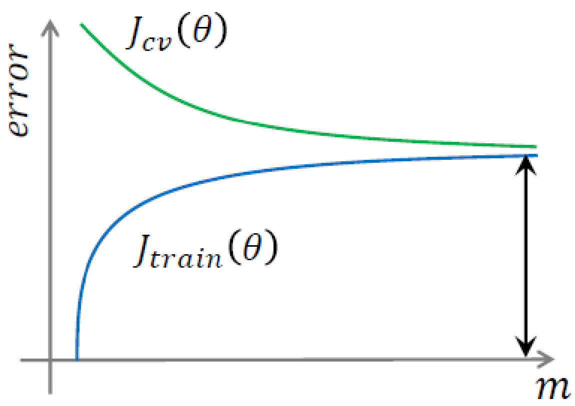


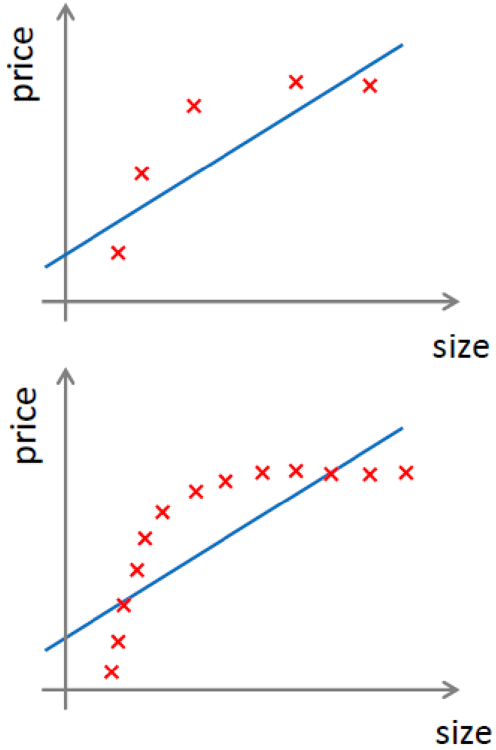
$$h_\theta(x) = \theta_0 + \theta_1 x + \theta_2 x^2$$



فصل هشتم : بایاس بالا

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

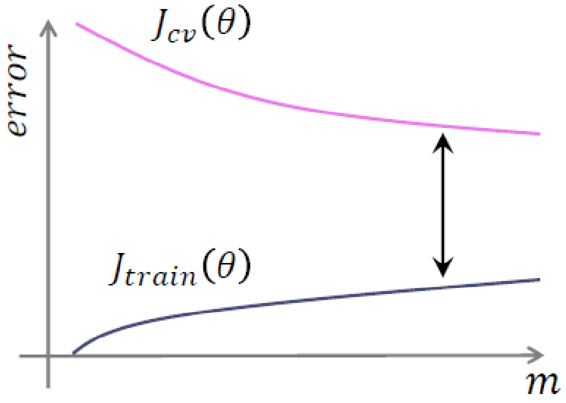


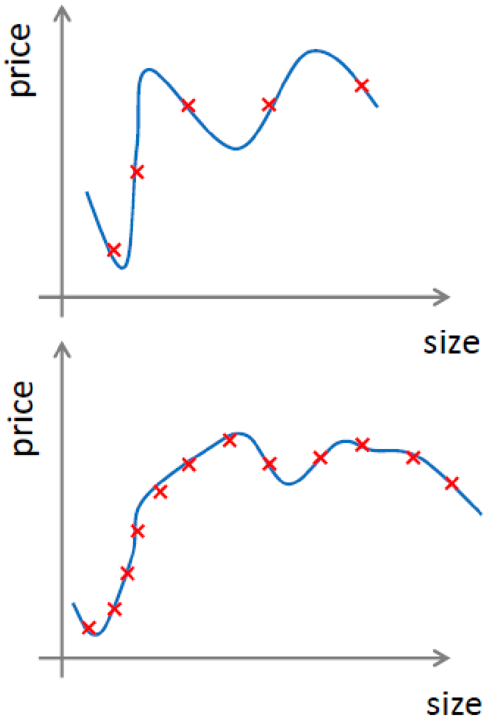


اگر یک الگوریتم یادگیری از بایاس بالا رنج ببرد، افزایش تعداد نمونه‌های آموزشی کمک چندانی به آن نخواهد کرد.

فصل هشتم : واریانس بالا

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

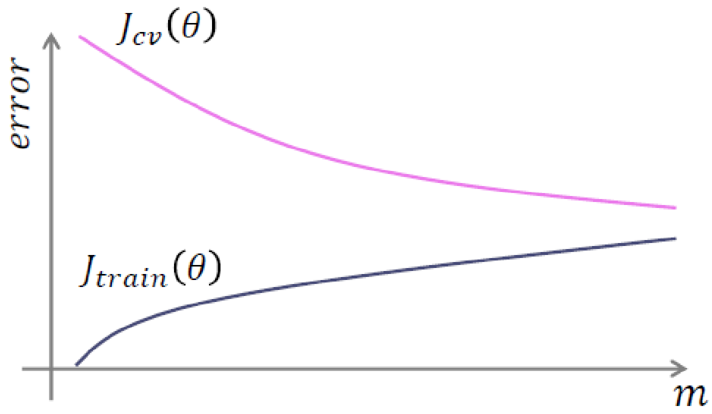


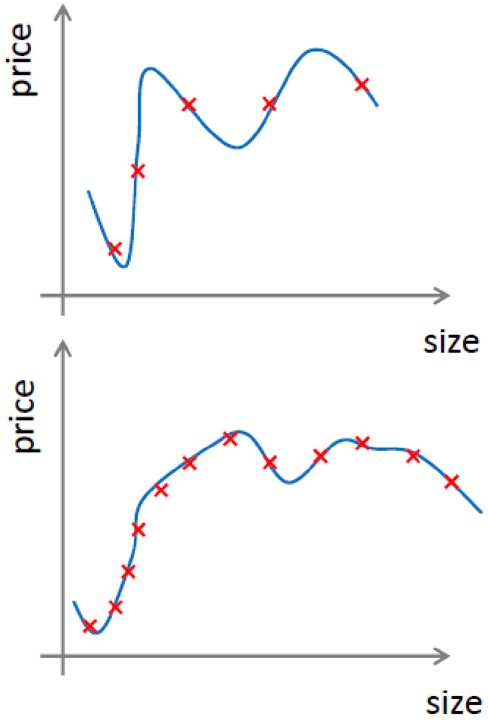


اگر یک الگوریتم یادگیری از واریانس بالا رنج ببرد، افزایش تعداد نمونه‌های آموزشی احتمالاً کمک کننده خواهد بود.

فصل هشتم : واریانس بالا

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





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

عیب‌یابی

فصل هشتم : عیب یابی

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

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

- 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؟

فصل هشتم : عیب یابی

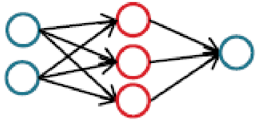
- راه‌حل‌های احتمالی.

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

فصل هشتم : شبکه های عصبی و بیش برازش

شبکه‌ی عصبی «کوچک»

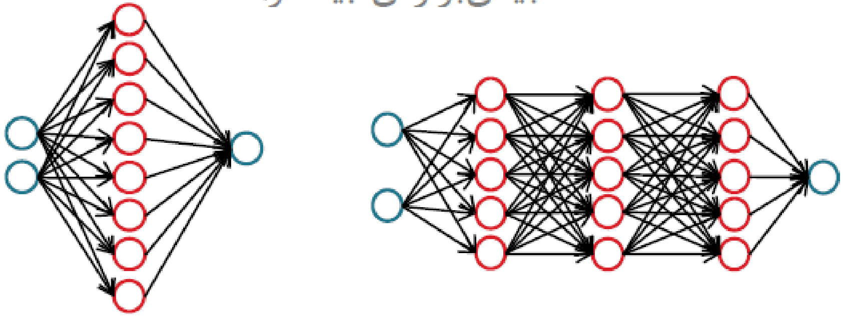
(پارامترهای کمتر؛ احتمال
کم‌برازش بیشتر)



هزینه محاسباتی کمتر

شبکه‌ی عصبی «بزرگ»

(پارامترهای بیشتر؛ احتمال
بیش‌برازش بیشتر)



هزینه محاسباتی بیشتر

استفاده از تنظیم برای برخورد با بیش‌برازش

با تشکر از توجه شما