سؤال مهم : آیا امکان دارد که یک loss در جایی کاهش یابد

که یک loss دیگر افزایش یابد؟

Ex: { MAE , MSE }

Error : $\underbrace{y_{true} - y_{predicted}}_{\propto \quad x}$

mAE / mse

$$mAE = \frac{1}{n} \sum |y_t - y_p|$$

$$mse = \frac{1}{n} \sum (y_t - y_p)^2$$

1

$y_t > y_p \longrightarrow |y_t - y_p| = (y_t - y_p)$

1 $mAE = \frac{1}{n} \sum (y_t - y_p) \rightsquigarrow mse = n\left(\frac{1}{n}\right)\left(\frac{1}{n}\right) \sum (y_t - y_p)(y_t - y_p)$

$mse = n\,(mAE)^2 \longrightarrow$

$$\boxed{mse = n(mae)^2}$$

α

$2uu'$

$$\frac{\partial}{\partial w} v^2 = 2v \frac{\partial v}{\partial w}$$

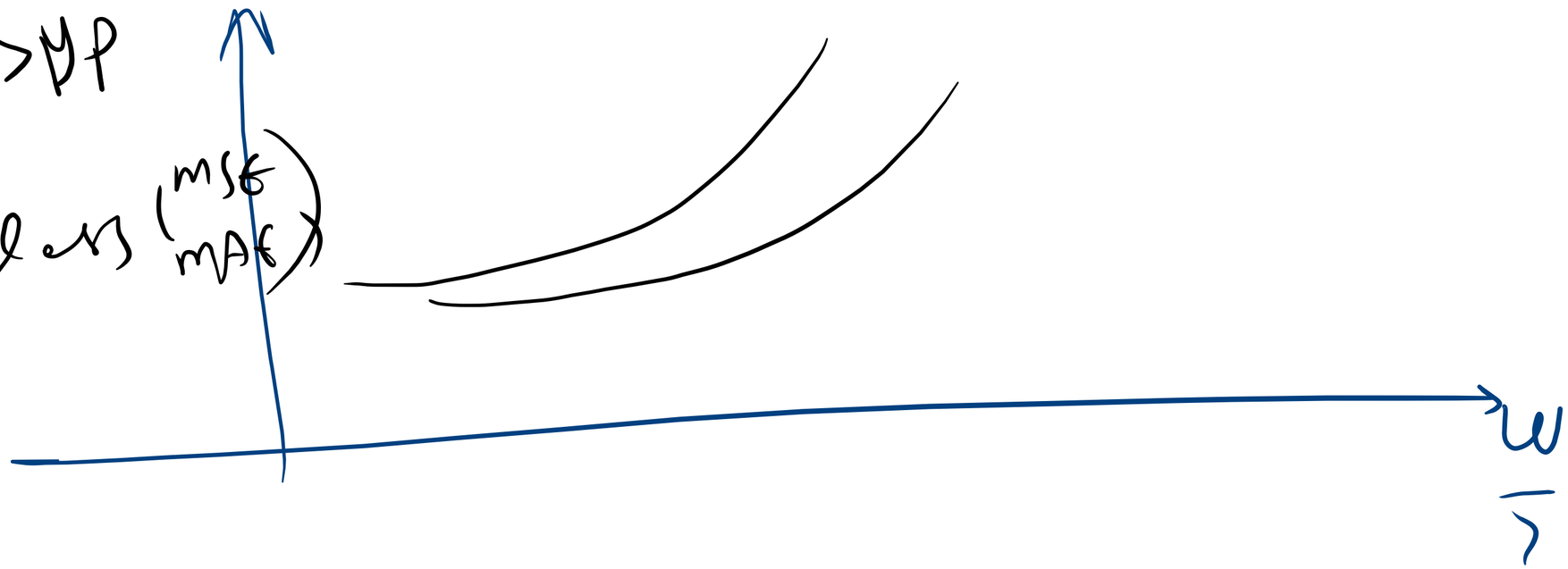
$$\frac{\partial mse}{\partial w} = n \frac{\partial}{\partial w} (mae)^2 \Rightarrow 2n(mae) \frac{\partial mae}{\partial w}$$

$y_t > y_p$

$$\frac{\partial mse}{\partial w} = 2n(mae) \frac{\partial mae}{\partial u}$$

$y_A > y_P$

loss (MSE, MAE)



2

$$y_t < y_p$$

$$y_t - y_p < 0$$

$$MSE = \frac{1}{n} \sum (y_t - y_p)^2$$

$$MAE = \frac{1}{n} \sum |y_t - y_p|$$

$$[y_t < y_p] \approx [y_t > y_p]$$

$y_t < y_p$

$$\begin{cases} MSE = \frac{1}{n}\sum \underline{(y_t - y_p)^2} \\ MAE = \frac{1}{n}\sum |y_t - y_p| \end{cases}$$

$$\begin{cases} (y_t - y_p)^2 = (y_p - y_t)^2 \\ |y_t - y_p| = |y_p - y_t| \end{cases}$$

$$[y_t < y_p] \approx [y_t < y_p]$$

③ $y_t = y_p$

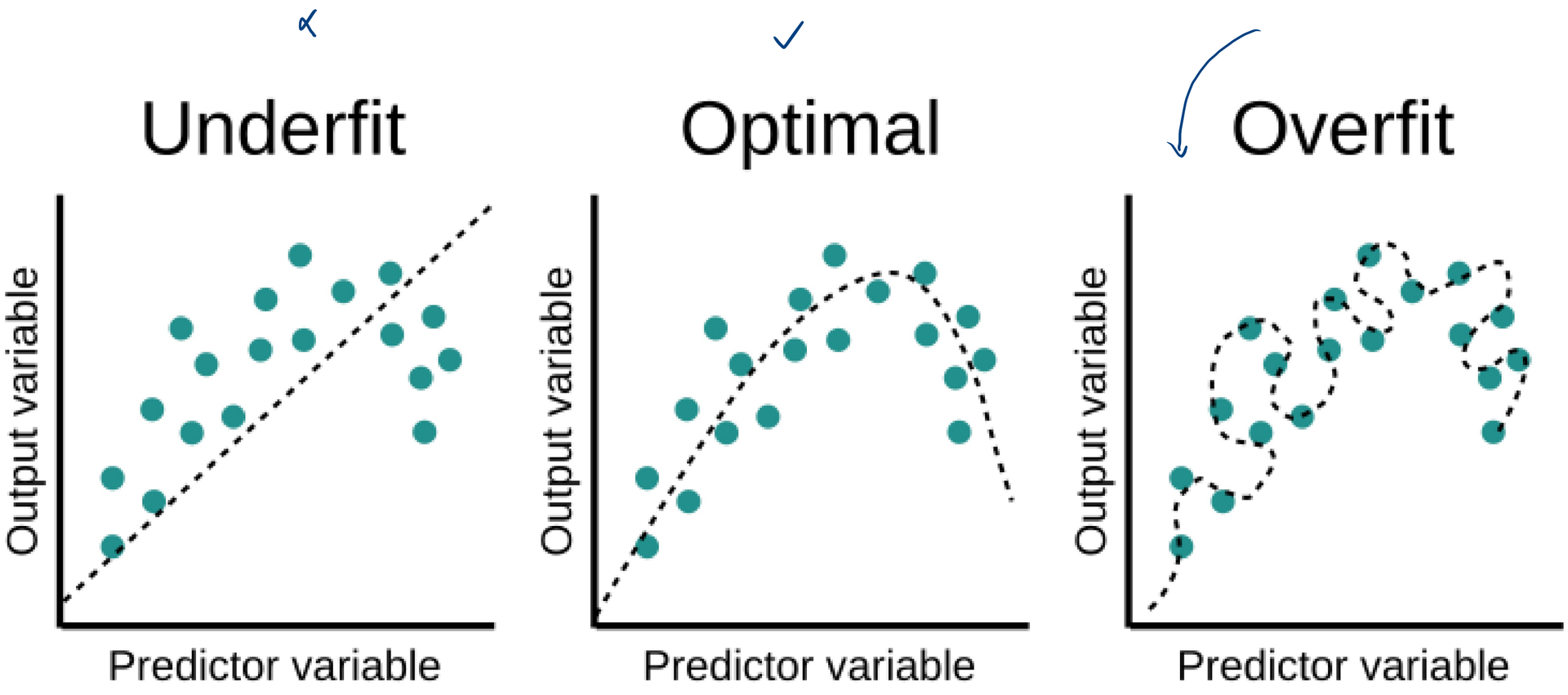
$$MSE = \frac{1}{n}\sum (y_t - y_p)^2 = 0$$

$$MAE = \frac{1}{n}\sum |y_t - y_p|^2 = 0$$

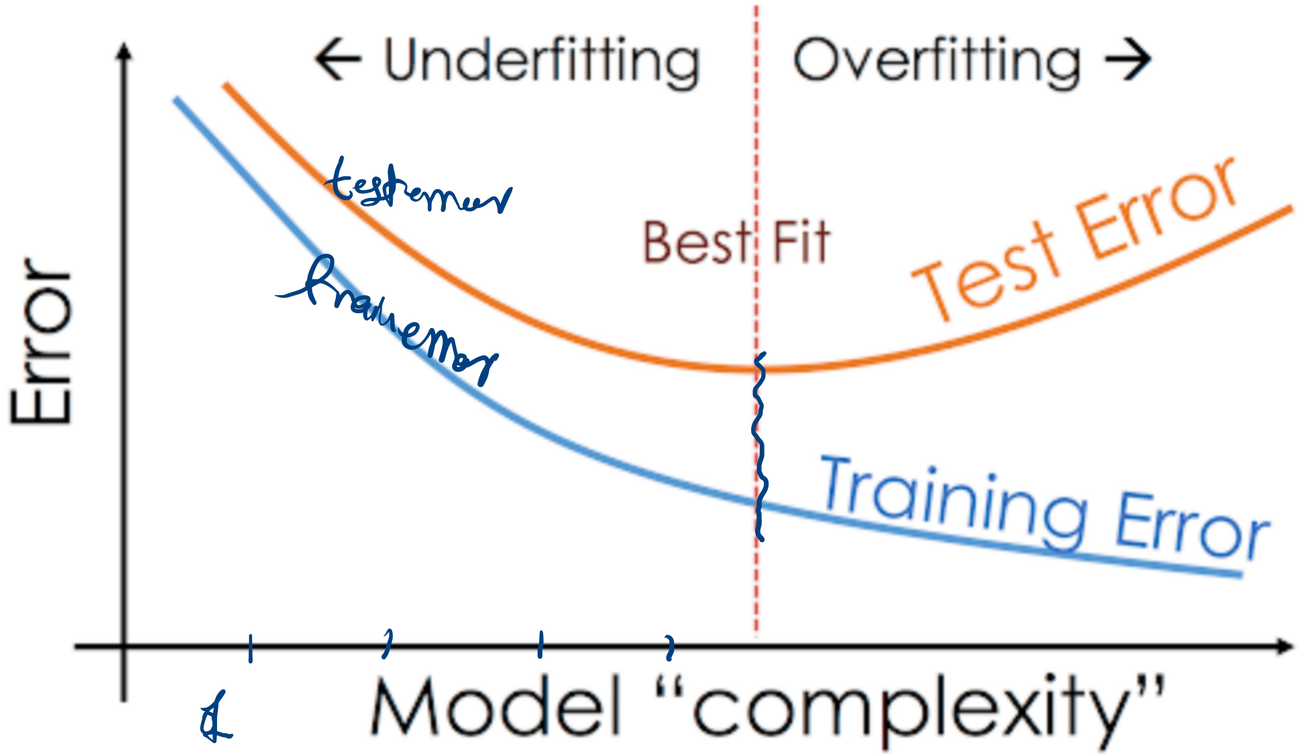
نحوه جلوگیری از بیش برازش و کم برازش

{Avoiding Under–fitting & Over–fitting }

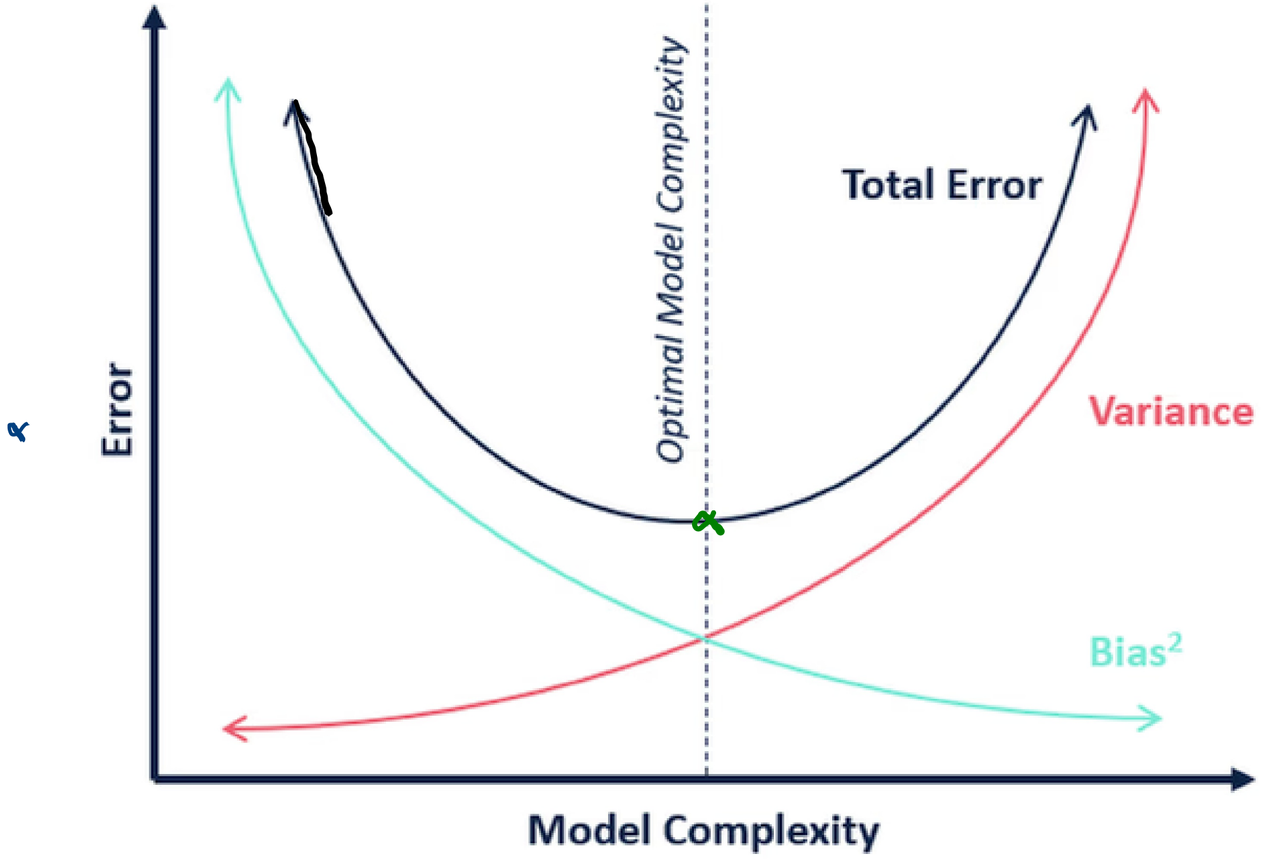
جلوگیری از Overfitting و Underfitting



جلوگیری از Overfitting و Underfitting



جلوگیری از Overfitting و Underfitting



تعداد صفحه کتاب (w_2 × X_2)	وزن کتاب (w_1 × X_1)	قیمت کتاب (y)
10,000	2	200
(100,000)	5	350
1,000,000	80	1500

باید نرمال شوند!

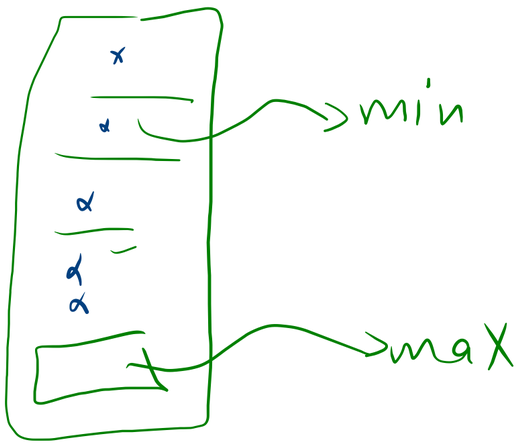
تعداد صفحه	وزن
0-1	0-1

$$y = w_1 (X_1) + w_2 X_2$$

$$y = w_1 + w_2 (10{,}000)$$

$$200 = \overset{0}{w_1} + \overset{\frac{1}{50}}{w_2} (10{,}000)$$

[عدم تاثیر تندلی غیرمنصفانه برای مدل]



۱. Normalization (نرمال‌سازی)

تبدیل مقادیر ویژگی‌ها به بازه‌ای ثابت (معمولاً بین 0 و 1)

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

train

Area	price
200	
300	
100	
50	

$$x = \frac{x - x_{min}}{x_{max} - x_{min}}$$

$x_{min} = 50$

$x_{max} = 300$

وضعنه

80 → $x_{min} = 50$, $x_{max} = 300$ → norm: → model.

min max

۰ ۲

جلوگیری از Overfitting و Underfitting : نرمال کردن

- **نرمال‌سازی باعث می‌شود همه ویژگی‌ها در مقیاس یکسان باشند.**

- فقط برای داده‌های پیوسته مناسب است.
- روی داده‌های گسسته (مثل جنسیت ۰ یا ۱) نباید انجام شود.
- در حضور داده‌های پرت (outliers) ممکن است دچار انحراف شود.

Minimum values of features

Maximum target range

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \times (max_{range} - min_{range}) + min_{range}$$

Maximum values of features

Minimum target range

$$X = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (2)$$

جلوگیری از Overfitting و Underfitting : استاندارد سازی

۲. Standardization **(استانداردسازی)**

- برای داده‌های پیوسته مناسب است.
- روی داده‌های گسسته یا باینری (مثل ۰/۱ یا دسته‌ای) نباید استفاده شود.
- در برابر داده‌های پرت نسبت به نرمال سازی مقاوم تر است اما همچنان حساس است!

$$z = \frac{x - \mu}{\sigma}$$

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}$$

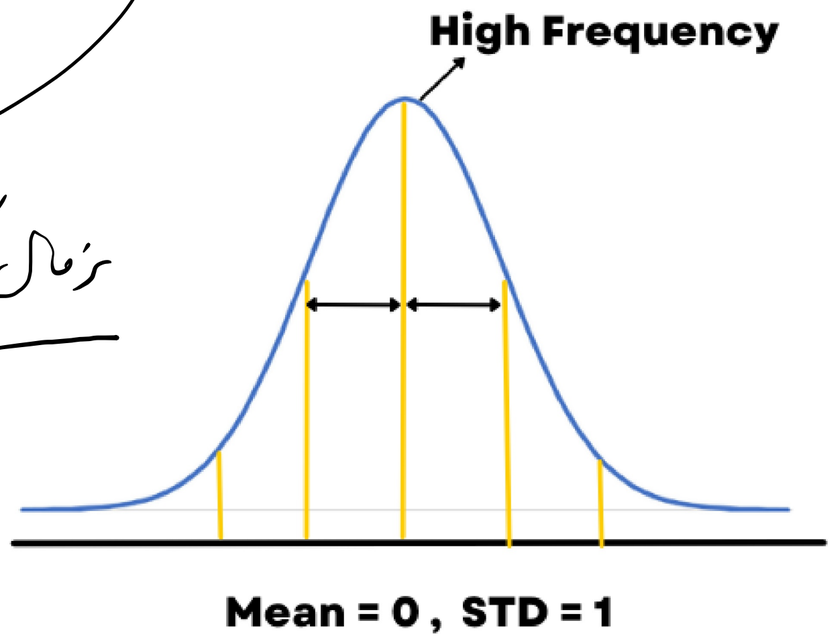
$$\mu = \frac{1}{N}\sum_{i=1}^{N}(x_i)$$

جلوگیری از Overfitting و Underfitting : استاندارد سازی

normal distribution

توزیع گاوسی

نرمال



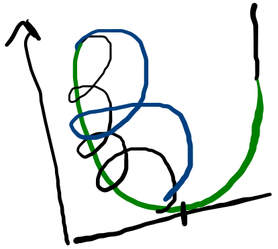
SVM

Regularization

α feature engineering

لندسی دیگر!

w_1	w_2	
X_1	X_2	y



mse

w_2 یک مقدار کمتر بدهیم w_1 حذف شود!

مدل با وزن‌های کمتر ما سریع‌تر همگرا می‌شود

less مدل

پارامتر!

۲. Regularization **(تنظیم مدل)**

- ایده‌ی اصلی: مجازات کردن وزن‌های بزرگ.

نوع ۱

L1 Regularization (Lasso)

α

جلوگیری از Overfitting و Underfitting : تنظیم مدل

x_1	x_2	x_3	y

hyperparam ← هزینه با متغیرک توسط هنر تقسیم می‌شود!

hp ←

$$J(w) = \frac{1}{2n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

$$J(w) = \underbrace{\frac{1}{2n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}_{\text{loss}} + \lambda \sum_{j=1}^{p}|w_j|$$

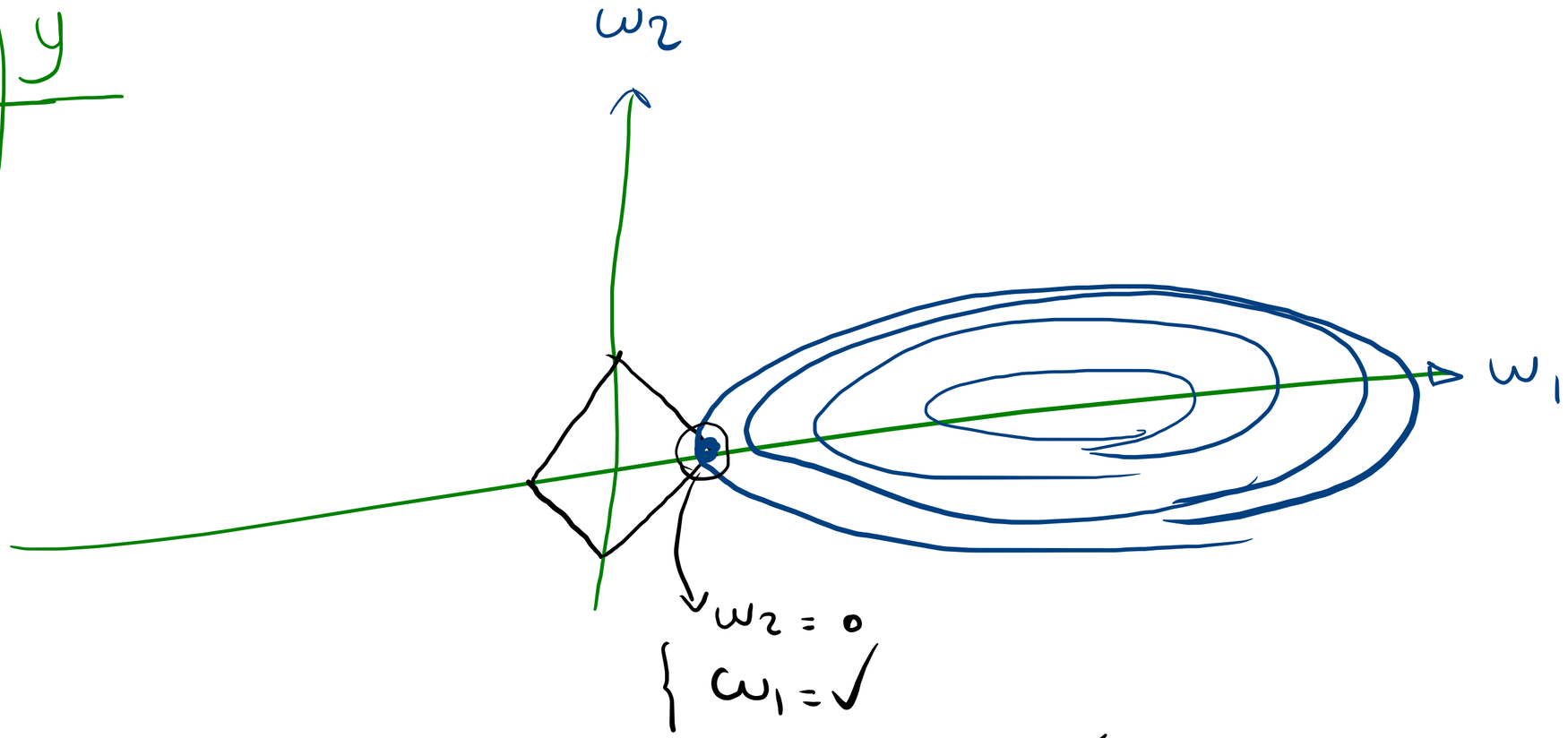
Penalty Term

$\lambda\left[|w_1| + |w_2| + |w_3| + \dots\right]$

w_1	w_2	
x_1	x_2	y

Lasso:

GD(mse)



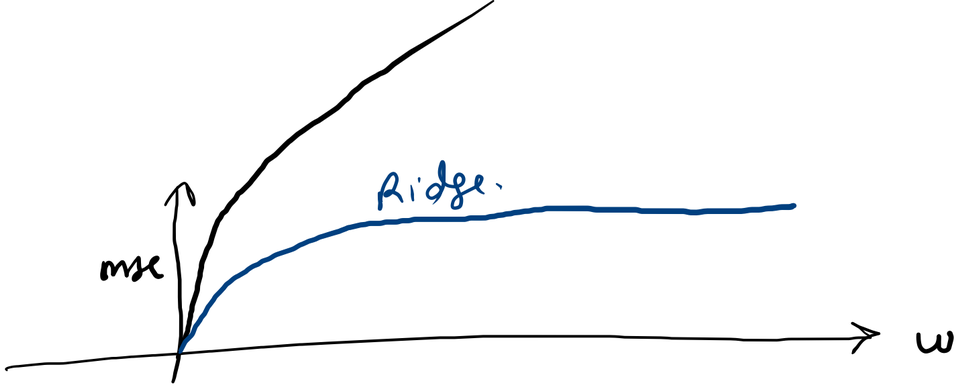
lasso ← من تضمین نمی‌کنم که برای $w_2 = 0$ ، w_1 دارای کمترین loss باشد.

اما تضمین دارم که برای یک loss معقول ، دچار overfitting هم نمی‌شوم!

نوع 2

L۲ Reg. (Ridge)

α



جلوگیری از Overfitting و Underfitting : تنظیم مدل

$$J(w) = \frac{1}{2n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

$$J(w) = \frac{1}{2n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^{p} w_j^2$$

Penalty Term

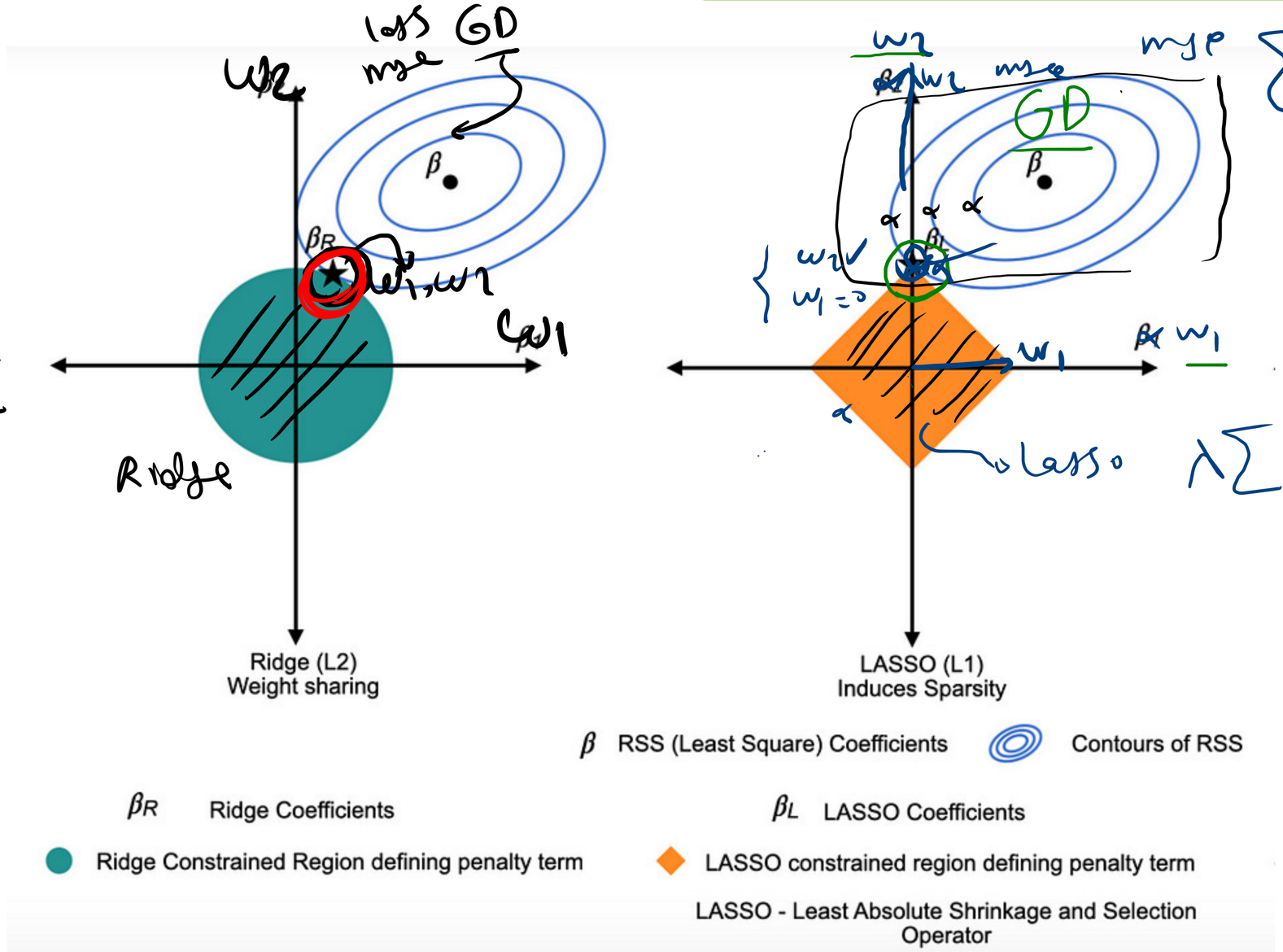
Ridge ← تنظیم مدل با کوچک نگه داشتن وزن‌ها در پیشگیری از overfitting کمک می‌کند

جلوگیری از Overfitting و Underfitting : تنظیم مدل

$\sum |w| \rightarrow |w_1| + |w_2| : \text{const}$

$\sum (w)^2 \rightarrow$

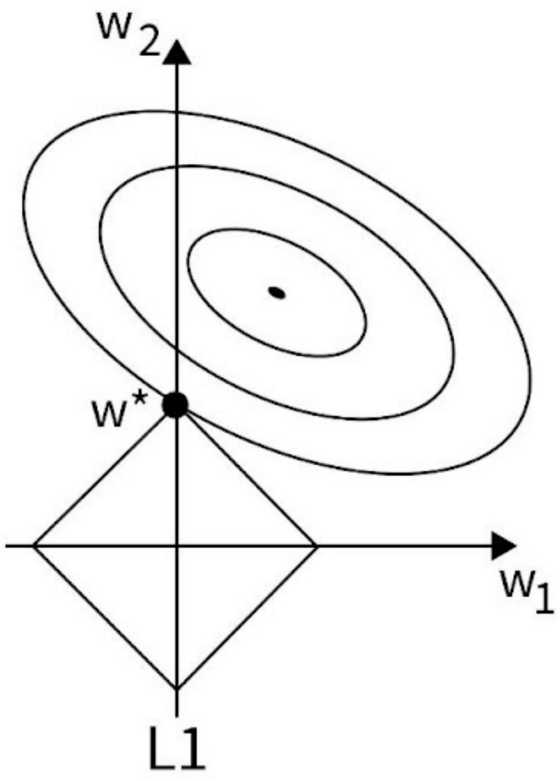
$w_1^2 + w_2^2 = \text{const}$

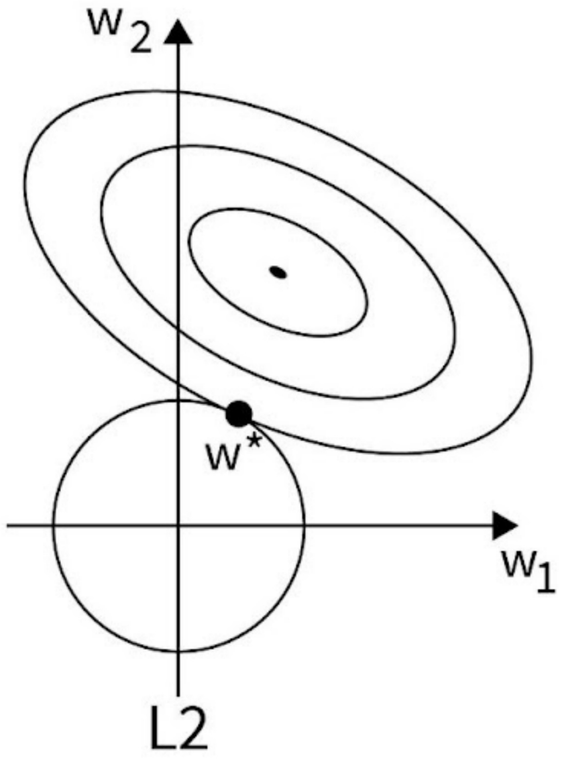


$\sum (y - y_p)^2$

$\lambda \sum |w_i|$

جلوگیری از Overfitting و Underfitting : تنظیم مدل





اگر داده‌ها پر از هم‌بستگی (correlation) بین ویژگی‌ها باشند،

Lasso ناپایدار است، ولی Ridge خوب عمل می‌کند.

اما Ridge همه ویژگی‌ها را نگه می‌دارد، در حالی‌که ما ممکن است

بخواهیم برخی حذف شوند.

نوع ۳

Elastic Net

ثبات و نرم بودن وزن‌ها (از L2) + انتخاب خودکار ویژگی‌ها (از L1)

mse

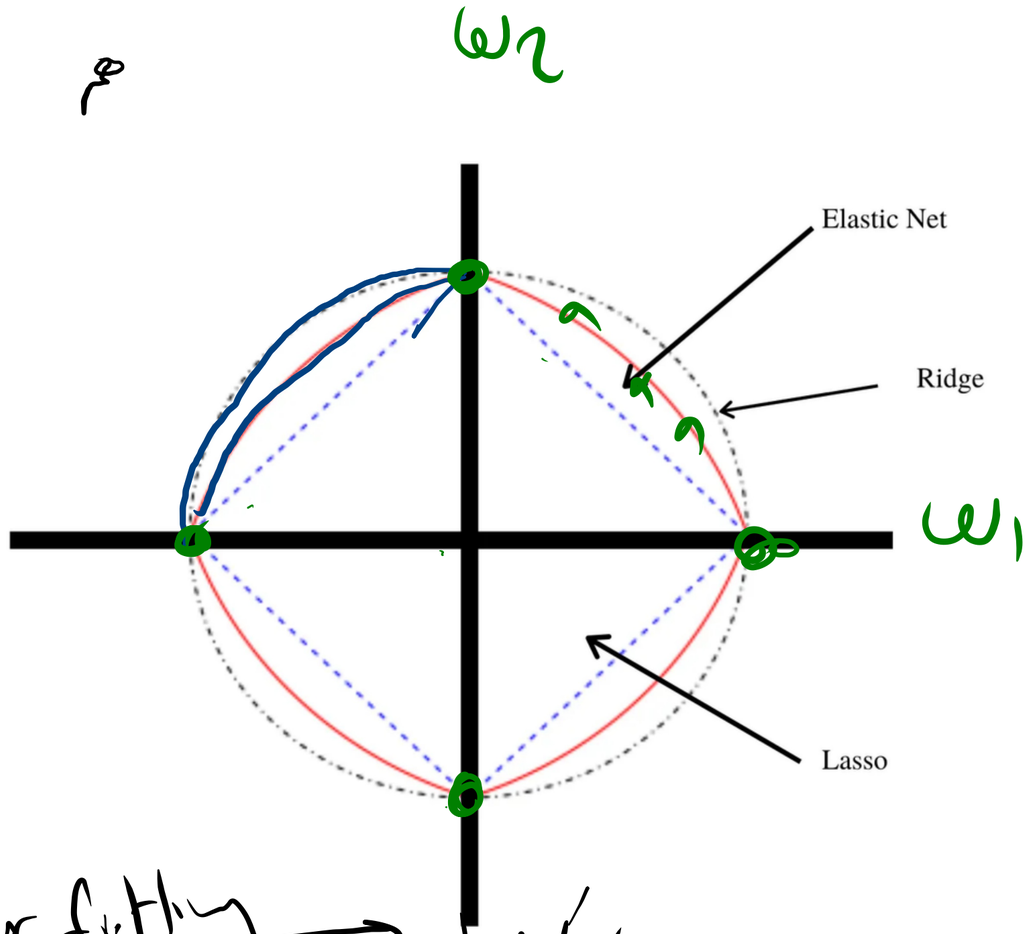
lasso

ridge.

$$J(w) = \frac{1}{2n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \lambda_1\|w\|_1 + \lambda_2\|w\|_2^2$$

Elastic : Ridge + Lasso

جلوگیری از Overfitting و Underfitting : تنظیم مدل



w_1 , w_2 هم

پارامترها

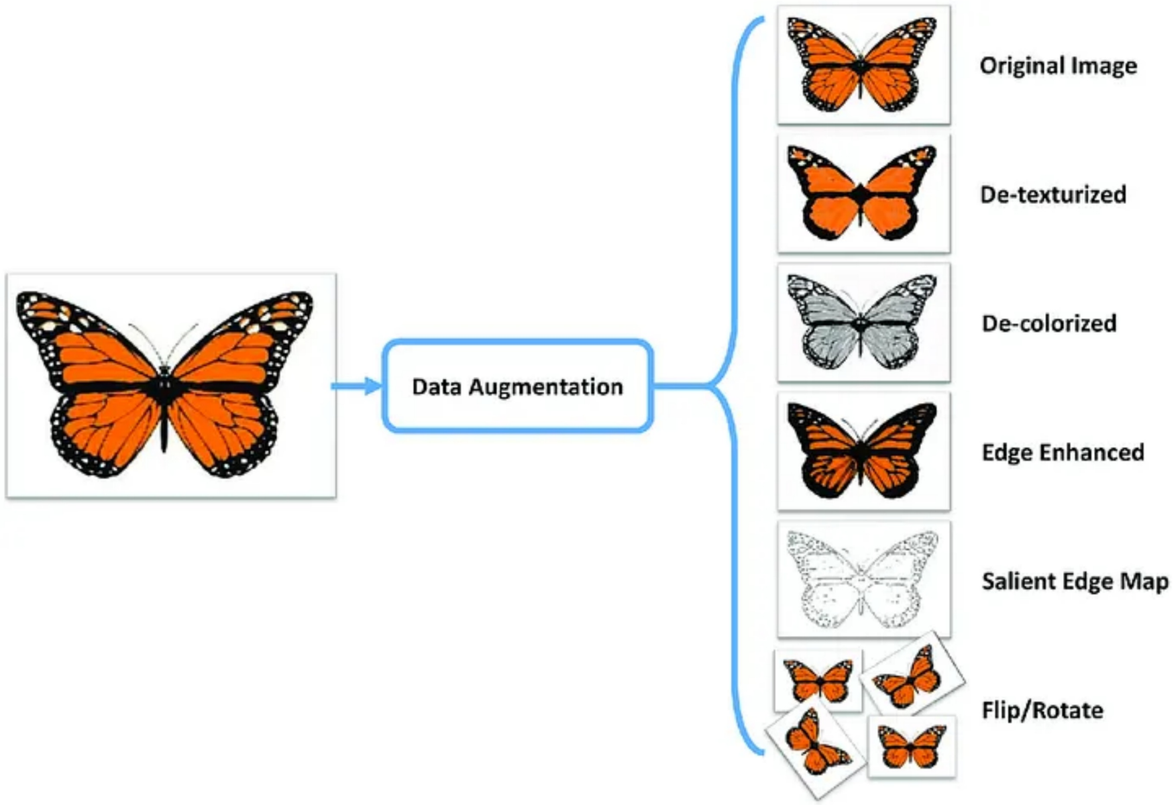
مقدار بزرگ

w_1 هم

آزادشان کنیم!

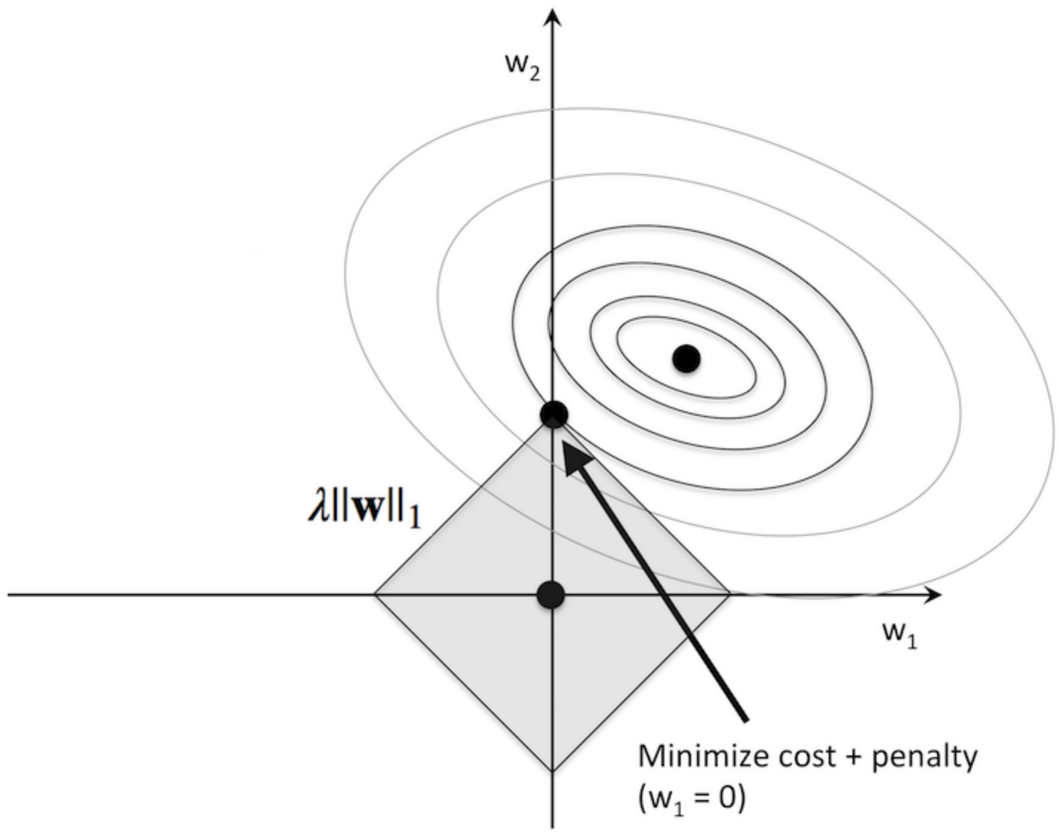
overfitting → مشکل!

- Data Augmentation



- PCA (Feature Eng.)

Next Sessions


w_2
$\lambda \|\mathbf{w}\|_1$
w_1
Minimize cost + penalty
($w_1 = 0$)