

رسالة محمد

منظم سازی

Regularization

# جریمه اندازه پارامترها

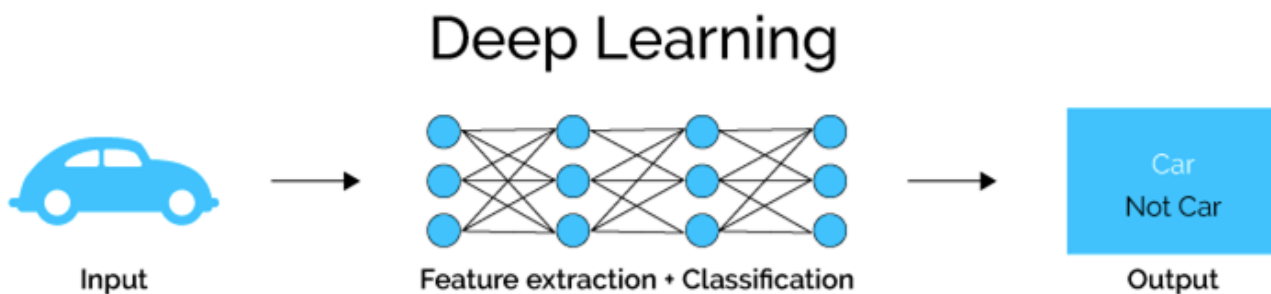
- به منظور محدود کردن ظرفیت یادگیری مدل، می‌توانیم محدودیت‌هایی بر روی پارامترهای مدل اعمال کنیم

$$\tilde{J}(\theta; X, y) = J(\theta; X, y) + \alpha \Omega(\theta)$$

- به طور معمول، محدودیتی روی بایاس‌ها انجام نمی‌شود بلکه ضرایب تبدیل Affine محدود می‌شوند

$$g(W^T x + b)$$

- ما از  $w$  برای مشخص کردن وزن‌هایی که از جریمه اندازه متاثر می‌شوند استفاده می‌کنیم و تمام پارامترهای شبکه را با  $\theta$  نمایش می‌دهیم

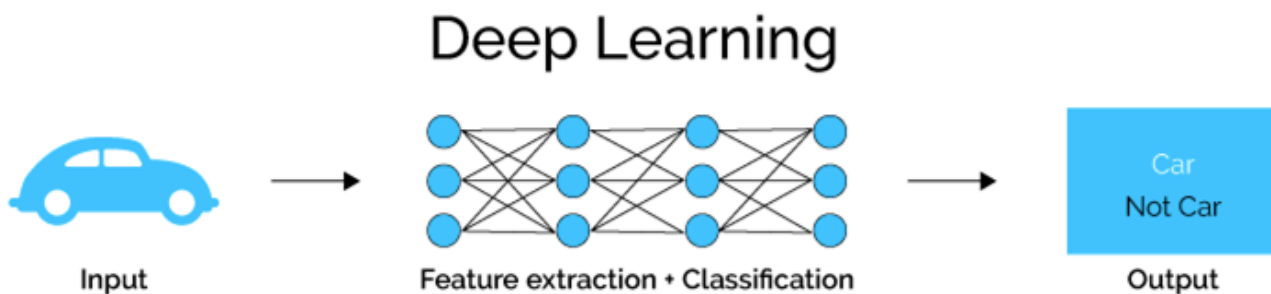


# جریمه اندازه پارامترها

- در برخی مواقع مناسب است تا برای هر لایه از یک جریمه جداگانه با ضریب  $\alpha$  متفاوت استفاده شود

$$\tilde{J}(\theta; X, y) = J(\theta; X, y) + \alpha \Omega(w)$$

- از آنجا که جستجو برای مقادیر مناسب چندین هایپرپارامتر می‌تواند هزینه زیادی داشته باشد، به طور معمول از یک ضریب یکسان در تمام لایه‌ها استفاده می‌شود



# منظم‌سازی پارامتر L2

- این استراتژی وزن‌های شبکه را به سمت مبدا نزدیک می‌کند

$$\Omega(\mathbf{w}) = \frac{1}{2} \|\mathbf{w}\|_2^2$$

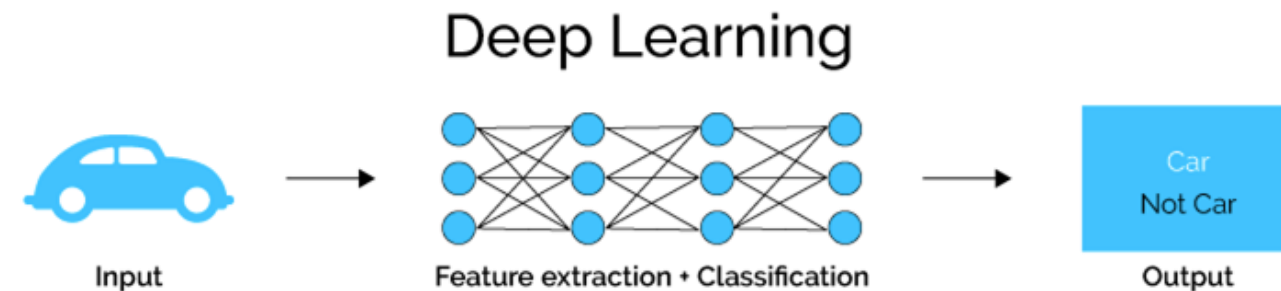
$$\tilde{J}(\boldsymbol{\theta}; X, \mathbf{y}) = \frac{\alpha}{2} \mathbf{w}^T \mathbf{w} + J(\boldsymbol{\theta}; X, \mathbf{y})$$

- گرادیان تابع ضرر منظم‌شده:

$$\nabla_{\mathbf{w}} \tilde{J}(\boldsymbol{\theta}; X, \mathbf{y}) = \alpha \mathbf{w} + \nabla_{\mathbf{w}} J(\boldsymbol{\theta}; X, \mathbf{y})$$

$$\mathbf{w} \leftarrow \mathbf{w} - \epsilon (\alpha \mathbf{w} + \nabla_{\mathbf{w}} J(\boldsymbol{\theta}; X, \mathbf{y}))$$

$$\mathbf{w} \leftarrow (1 - \epsilon \alpha) \mathbf{w} - \epsilon \nabla_{\mathbf{w}} J(\boldsymbol{\theta}; X, \mathbf{y})$$

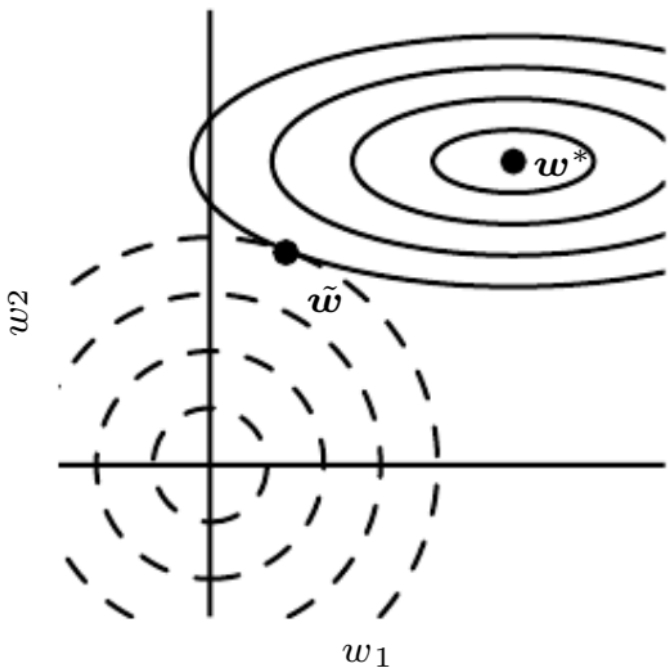


# منظم سازی پارامتر L2

$$\Omega(\mathbf{w}) = \frac{1}{2} \|\mathbf{w}\|_2^2$$

$$\mathbf{w} \leftarrow (1 - \epsilon\alpha)\mathbf{w} - \epsilon \nabla_{\mathbf{w}} J(\boldsymbol{\theta}; X, \mathbf{y})$$

- قبل از انجام به روزرسانی معمولی مبتنی بر گرادیان، بردار وزن را در هر گام با یک ضریب ثابت کاهش می دهد
- وزن هایی که تغییر کمتری در تابع ضرر ایجاد می کنند، اهمیت کمتری دارند و بیشتر کاهش می یابند



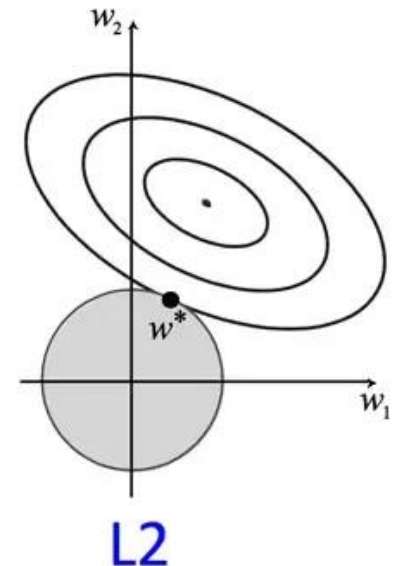
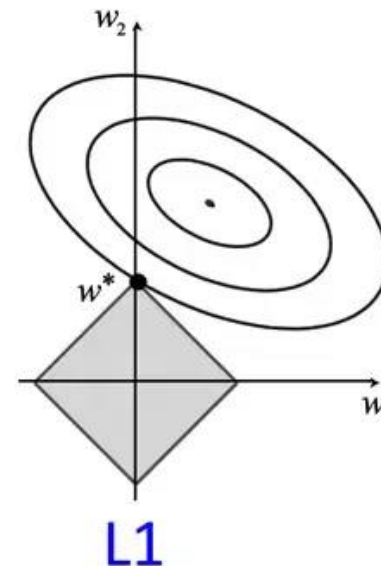
# منظم‌سازی پارامتر L1

- منظم‌سازی L1 بر روی پارامترهای شبکه  $\mathbf{w}$  به صورت زیر تعریف می‌شود

$$\Omega(\mathbf{w}) = \|\mathbf{w}\|_1 = \sum_i |w_i|$$

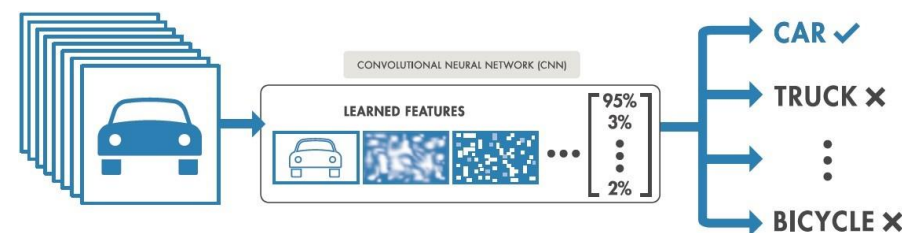
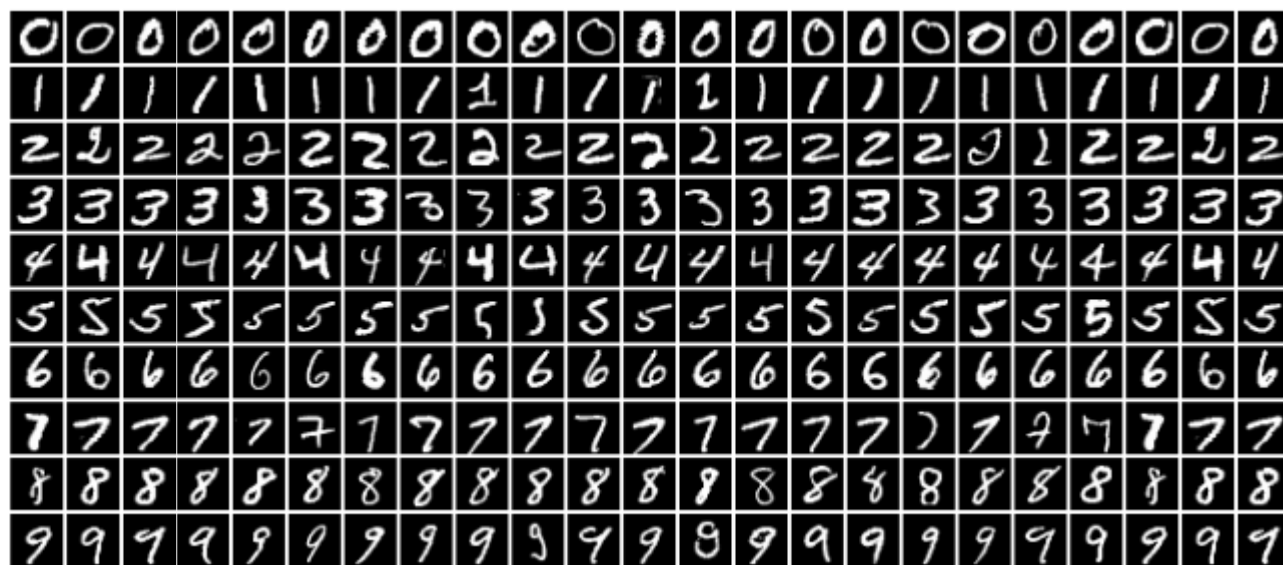
$$\tilde{J}(\boldsymbol{\theta}; \mathbf{X}, \mathbf{y}) = \alpha \|\mathbf{w}\|_1 + J(\boldsymbol{\theta}; \mathbf{X}, \mathbf{y})$$

$$\nabla_{\mathbf{w}} \tilde{J}(\boldsymbol{\theta}; \mathbf{X}, \mathbf{y}) = \alpha \text{sign}(\mathbf{w}) + \nabla_{\mathbf{w}} J(\boldsymbol{\theta}; \mathbf{X}, \mathbf{y})$$

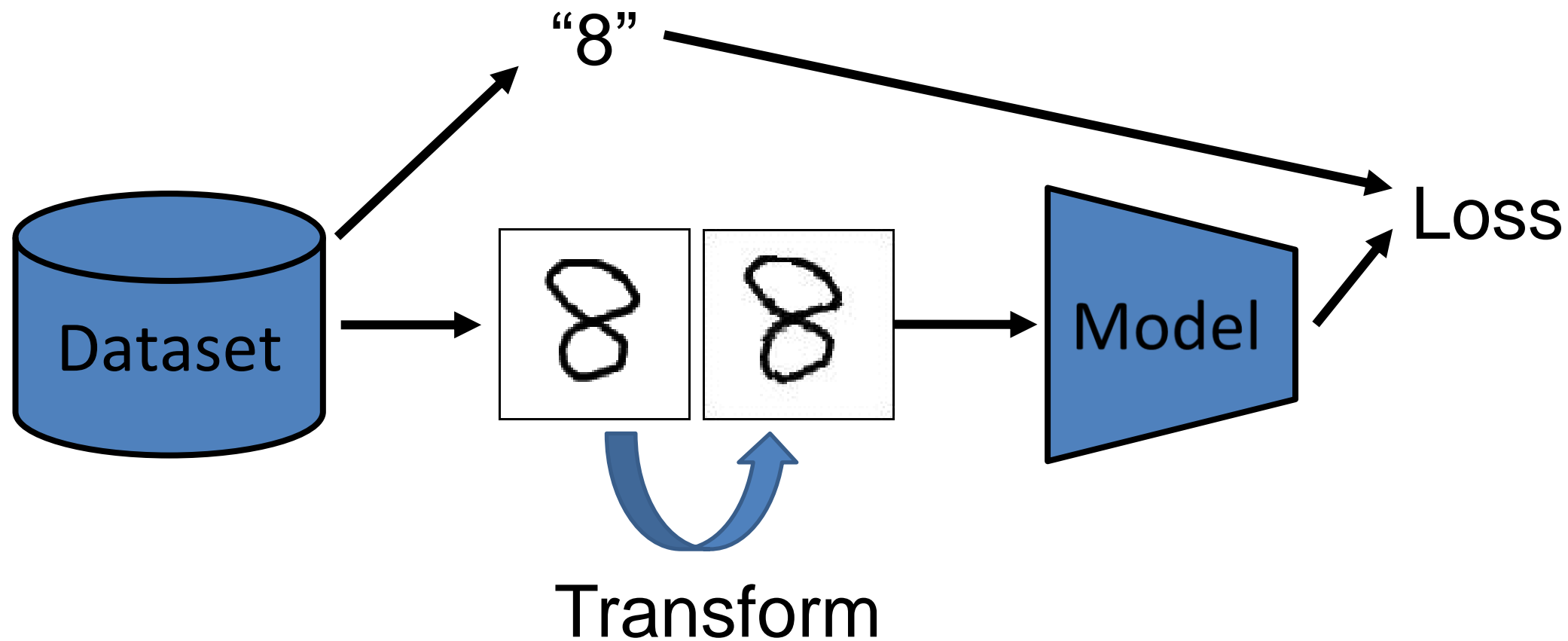


# داده‌افزایی (Augmentation)

- بهترین راه برای افزایش قدرت تعمیم‌دهی یک الگوریتم یادگیری ماشین با ظرفیت یادگیری بالا، آموزش آن بر روی داده‌های بیشتر است
- جمع‌آوری داده معمولاً فرآیند دشوار و خسته‌کننده‌ای است
- می‌توانیم داده‌های ساختگی بسازیم و به داده‌های آموزشی اضافه کنیم
- این کار برای مسئله دسته‌بندی راحت‌ترین است
- می‌توانیم جفت‌های  $(x, y)$  جدید را به سادگی و تنها با تبدیل  $x$  بسازیم







# داده‌افزایی: flip



b

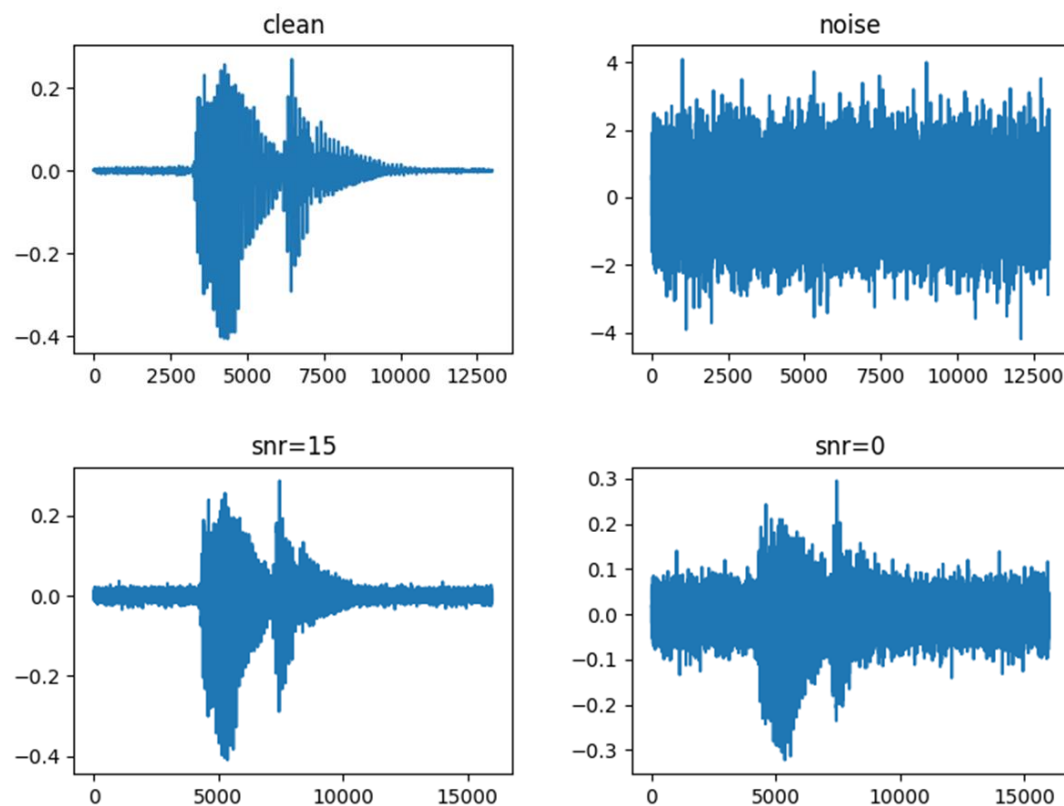
d

p

q

# داده‌افزایی: افزودن نویز

- افزودن نویز در ورودی به یک شبکه عصبی می‌تواند به عنوان نوعی داده‌افزایی در نظر گرفته شود



- برای بسیاری از مسائل دسته‌بندی و حتی برخی از مسائل رگرسیون، با افزودن مقدار محدودی نویز به ورودی همچنان می‌توان همان خروجی را توقع داشت
- افزودن نویز می‌تواند در لایه‌های میانی شبکه نیز انجام شود