توابع هزینه و بهینه‌سازی

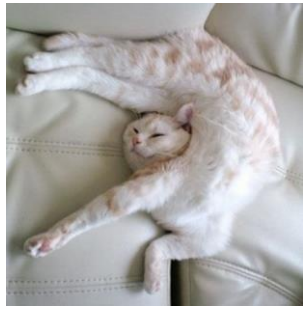
سید ناصر رضوی www.snrazavi.ir

۱۳۹۷

یادآوری: چالش‌های دسته‌بندی تصاویر

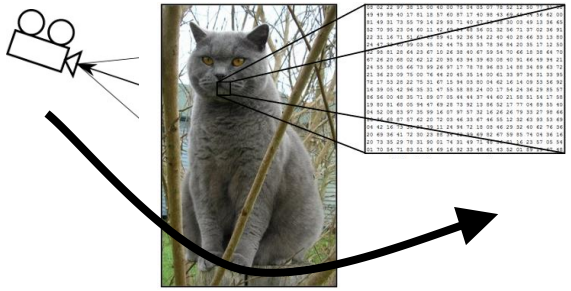
انسداد

تغییر شکل



نورپردازی

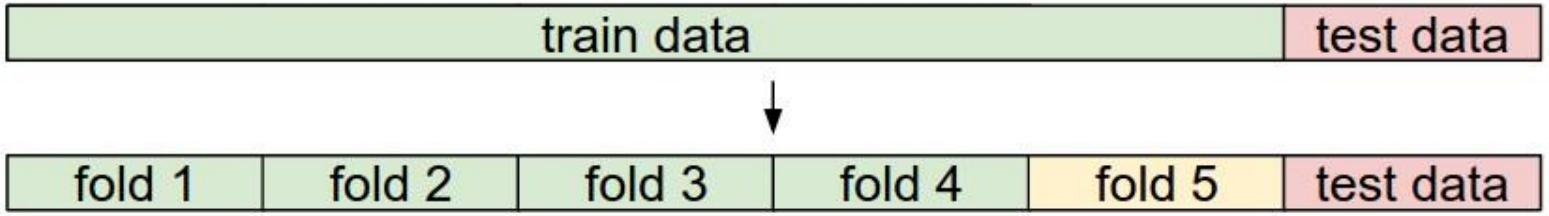
زاویه‌ی دوربین

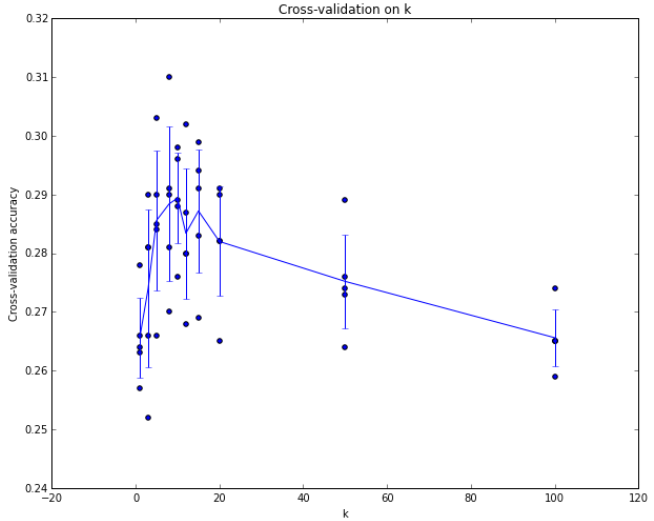


تنوع درون کلاسی

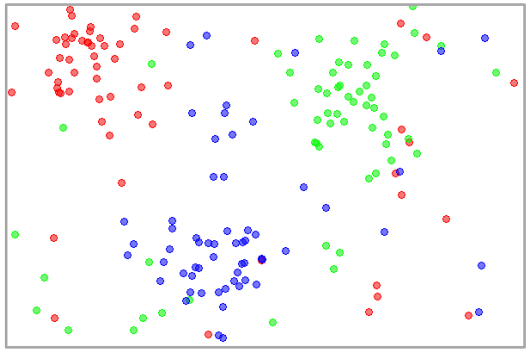
ترکیب با پس‌زمینه

یادآوری: رویکرد مبتنی بر داده، kNN

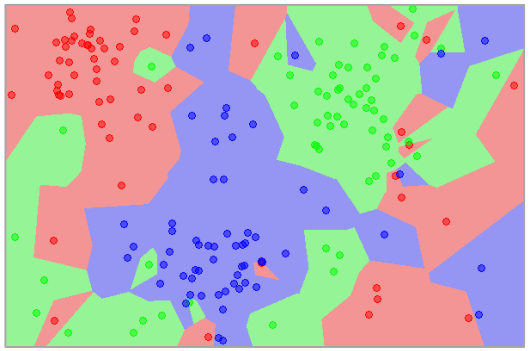




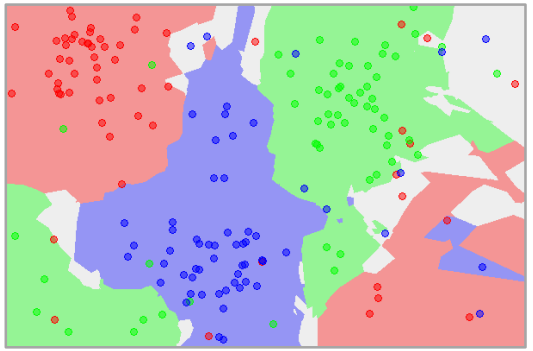
داده‌ها



نزدیک‌ترین همسایه



۵-نزدیک‌ترین همسایه



تابع هزینه و بهینه‌سازی

- تعریف تابع هزینه.
  - به گونه‌ای که بیانگر میزان ناخشنودی ما از امتیازهای محاسبه شده بر روی داده‌های مجموعه آموزشی باشد.

- بهینه‌سازی.
  - یافتن مقادیر پارامترها به گونه‌ای که تابع هزینه کمینه گردد.

هواپیما	-3.45	-0.51	3.42
خودرو	-8.87	**6.04**	4.64
پرنده	0.09	5.31	2.65
گربه	**2.90**	-4.22	5.10
گوزن	4.48	-4.19	2.64
سگ	**8.02**	3.58	5.55
قورباغه	3.78	4.49	**-4.34**
اسب	1.06	-4.37	-1.50
کشتی	-0.36	-2.09	-4.79
کامیون	-0.72	-2.93	**6.14**

تابع هزینه و بهینه‌سازی

- مجموعه آموزشی.
  - شامل ۳ نمونه آموزشی از ۳ کلاس مختلف.

گربه	**3.2**	1.3	2.2
خودرو	5.1	**4.9**	2.5
قورباغه	-1.7	2.0	**-3.1**

- تابع امتیاز.
  - محاسبه امتیاز هر کلاس برای هر یک از داده‌ها.

$$s = f(W, x) = Wx$$

تابع هزینه SVM چند کلاسی

- محاسبه هزینه برای داده (x_i, y_i)
  - x_i: تصویر ورودی
  - y_i: برچسب تصویر ورودی (یک عدد صحیح)

$$L_i = \sum_{j \neq y_i} \max(0, s_j - s_{y_i} + 1)$$



گربه	**3.2**	1.3	2.2
خودرو	5.1	**4.9**	2.5
قورباغه	-1.7	2.0	**-3.1**

تابع هزینه SVM چند کلاسی

- محاسبه هزینه برای داده (x_i, y_i)
  - x_i: تصویر ورودی
  - y_i: برچسب تصویر ورودی (یک عدد صحیح)

گربه	**3.2**	1.3	2.2
خودرو	5.1	**4.9**	2.5
قورباغه	-1.7	2.0	**-3.1**
	2.9		

$$L_i = \sum_{j \neq y_i} \max(0, s_j - s_{y_i} + 1)$$

```
= max(0,  5.1 - 3.2 + 1) +
  max(0, -1.7 - 3.2 + 1)
= max(0, 2.9) + max(0, -3.9)
= 2.9 + 0
= 2.9
```

تابع هزینه SVM چند کلاسی

- محاسبه هزینه برای داده (x_i, y_i)
  - x_i: تصویر ورودی
  - y_i: برچسب تصویر ورودی (یک عدد صحیح)



گربه	**3.2**	1.3	2.2
خودرو	5.1	**4.9**	2.5
قورباغه	-1.7	2.0	**-3.1**
	2.9	**0.0**	

$$L_i = \sum_{j \neq y_i} \max(0, s_j - s_{y_i} + 1)$$

```
= max(0, 1.3 - 4.9 + 1) +
  max(0, 2.0 - 4.9 + 1)
= max(0, -2.6) + max(0, -1.9)
= 0 + 0
= 0
```

تابع هزینه SVM چند کلاسی

- محاسبه هزینه برای داده (x_i, y_i)
  - x_i: تصویر ورودی
  - y_i: برچسب تصویر ورودی (یک عدد صحیح)

گربه	**3.2**	1.3	2.2
خودرو	5.1	**4.9**	2.5
قورباغه	-1.7	2.0	**-3.1**
	2.9	**0.0**	**12.9**

$$L_i = \sum_{j \neq y_i} \max(0, s_j - s_{y_i} + 1)$$

```
= max(0, 2.2 - (-3.1) + 1) +
  max(0, 2.5 - (-3.1) + 1)
= max(0, 6.3) + max(0, 6.6)
= 6.3 + 6.6
= 12.9
```

تابع هزینه SVM چند کلاسی

- محاسبه هزینه برای داده (x_i,y_i)
  - x_i: تصویر ورودی
  - y_i: برچسب تصویر ورودی (یک عدد صحیح)

$$L_i = \sum_{j \neq y_i} \max(0, s_j - s_{y_i} + 1)$$

- هزینه کل.
  - میانگین هزینه بر روی تمام داده‌های آموزشی.

$$L = \frac{1}{N} \sum_i L_i$$



گربه	**3.2**	1.3	2.2
خودرو	5.1	**4.9**	2.5
قورباغه	-1.7	2.0	**-3.1**
	2.9	**0.0**	**12.9**

L = (2.9 + 0 + 12.9) / 3 = 5.27

تابع هزینه SVM چند کلاسی

- محاسبه هزینه برای داده (x_i,y_i)
  - x_i: تصویر ورودی
  - y_i: برچسب تصویر ورودی (یک عدد صحیح)

$$L_i = \sum_{j \neq y_i} \max(0, s_j - s_{y_i} + 1)$$

گربه	**3.2**	1.3	2.2
خودرو	5.1	**4.9**	2.5
قورباغه	-1.7	2.0	**-3.1**
	2.9	0.0	12.9

اگر در رابطه بالا، مجموع روی تمام کلاس‌ها محاسبه شود، تابع هزینه چگونه تغییر می‌کند؟

تابع هزینه SVM چند کلاسی

- محاسبه هزینه برای داده (x_i, y_i)
  - x_i: تصویر ورودی
  - y_i: برچسب تصویر ورودی (یک عدد صحیح)

$$L_i = \sum_{j \neq y_i} \max(0, s_j - s_{y_i} + 1)$$

گربه	**3.2**	1.3	2.2
خودرو	5.1	**4.9**	2.5
قورباغه	-1.7	2.0	**-3.1**
	2.9	0.0	12.9

اگر در رابطه بالا، به جای مجموع از میانگین استفاده شود، تابع هزینه چگونه تغییر می‌کند؟

تابع هزینه SVM چند کلاسی

- محاسبه هزینه برای داده (x_i,y_i)
  - x_i: تصویر ورودی
  - y_i: برچسب تصویر ورودی (یک عدد صحیح)

گربه	**3.2**	1.3	2.2
خودرو	5.1	**4.9**	2.5
قورباغه	-1.7	2.0	**-3.1**
	2.9	0.0	12.9

$$L_i = \sum_{j \neq y_i} \max(0, s_j - s_{y_i} + 1)$$

اگر به جای رابطه بالا، از رابطه زیر استفاده شود، تابع هزینه چگونه تغییر می‌کند؟

$$L_i = \sum_{j \neq y_i} \max(0, s_j - s_{y_i} + 1)^2$$

تابع هزینه SVM چند کلاسی

- محاسبه هزینه برای داده‌ی (x_i,y_i)
  - x_i: تصویر ورودی
  - y_i: برچسب تصویر ورودی (یک عدد صحیح)

$$L_i = \sum_{j \neq y_i} \max(0, s_j - s_{y_i} + 1)$$

کمترین و بیشترین مقدار ممکن برای تابع هزینه چقدر است؟



گربه	**3.2**	1.3	2.2
خودرو	5.1	**4.9**	2.5
قورباغه	-1.7	2.0	**-3.1**
	2.9	0.0	12.9

تابع هزینه SVM چند کلاسی

- محاسبه هزینه برای داده (x_i,y_i)
  - x_i: تصویر ورودی
  - y_i: برچسب تصویر ورودی (یک عدد صحیح)

$$L_i = \sum_{j \neq y_i} \max(0, s_j - s_{y_i} + 1)$$



گربه	**3.2**	1.3	2.2
خودرو	5.1	**4.9**	2.5
قورباغه	-1.7	2.0	**-3.1**
	2.9	0.0	12.9

به طور معمول در لحظه شروع، مقادیر w با مقادیر تصادفی کوچک مقداردهی می‌شوند و بنابراین تمام امتیازها تقریبا صفر هستند. در این حالت مقدار تابع هزینه چقدر است؟

تابع هزینه SVM چند کلاسی

- پیاده‌سازی در پایتون.

$$L_i = \sum_{j \neq y_i} \max(0, s_j - s_{y_i} + 1)$$

```
def L_i_vectorized(x, y, w):
    scores = w.dot(x)
    margins = np.maximum(0, scores - scores[y] + 1)
    margins[y] = 0
    loss_i = np.sum(margins)
    return loss_i
```

scores - scores[y] + 1

3.2		3.2		1.0
5.1	−	3.2	+	1.0
-1.7		3.2		1.0

وجود مشکل در تابع هزینه

- فرض کنید یک W پیدا کرده‌ایم به طوری که $L = 0$. آیا W یکتا است؟

$$f(x, W) = Wx$$

$$L = \frac{1}{N} \sum_{i=1}^{N} \sum_{j \neq y_i} \max(0, f(x_i; W)_j - f(x_i; W)_{y_i} + 1)$$

وجود مشکل در تابع هزینه

$$L_i = \sum_{j \neq y_i} \max(0, s_j - s_{y_i} + 1)$$

گربه	**3.2**	1.3	2.2
خودرو	5.1	**4.9**	2.5
قورباغه	-1.7	2.0	**-3.1**
	2.9	0.0	

قبلاً:

```
= max(0, 1.3 - 4.9 + 1) +
  max(0, 2.0 - 4.9 + 1)
= max(0, -2.6) + max(0, -1.9)
= 0 + 0
= 0
```

اگر مقدار W را ۲ برابر کنیم:

```
= max(0, 2.6 - 9.8 + 1) +
  max(0, 4.0 - 9.8 + 1)
= max(0, -6.2) + max(0, -4.8)
= 0 + 0
= 0
```

راه‌حل: تنظیم وزن‌ها

- تنظیم وزن‌ها.

ضریب تنظیم

$$L = \frac{1}{N}\sum_{i=1}^{N}\sum_{j\neq y_i}\max(0, f(x_i;W)_j - f(x_i;W)_{y_i} + 1) + \boxed{\lambda R(W)}$$

- چند روش متداول.

$$R(W) = \sum_k \sum_l W_{k,l}^2$$ L2-Regularization

$$R(W) = \sum_k \sum_l |W_{k,l}|$$ L1-Regularization

$$R(W) = \sum_k \sum_l \beta W_{k,l}^2 + |W_{k,l}|$$ Elastic net (L1 + L2)

دسته‌بند سافت‌مکس (رگرسیون لجستیک چند کلاسی)

- امتیازها. لگاریتم احتمال کلاس‌ها به صورت نرمال نشده!

$$P(Y=k|X=x_i)=\frac{e^{s_k}}{\sum_j e^{s_j}}$$

تابع سافت‌مکس

$$s=f(x_i;W)$$

گربه	3.2
خودرو	5.1
قورباغه	-1.7

- هدف.
  - بیشینه‌سازی لگاریتم درست‌نمایی (یا کمینه‌سازی منفی لگاریتم درست‌نمایی)!

$$L_i=-\log P(Y=y_i|X=x_i)=-\log\left(\frac{e^{s_{y_i}}}{\sum_j e^{s_j}}\right)$$

دسته‌بند سافت‌مکس (رگرسیون لجستیک چند کلاسی)

- امتیازها. لگاریتم احتمال کلاس‌ها به صورت نرمال نشده!

$$L_i = -\log\left(\frac{e^{s_{y_i}}}{\sum_j e^{s_j}}\right)$$

$$L_i = -\log(0.13) = \mathbf{0.89}$$

	لگاریتم احتمال‌ها به صورت غیرنرمال		احتمال‌ها به صورت غیرنرمال		احتمال‌ها
گربه	3.2	توان رسانی →	24.5	نرمال‌سازی →	0.13
خودرو	5.1		164.0		0.87
قورباغه	-1.7		0.2		0.0

کمترین و بیشترین مقدار ممکن برای تابع هزینه چقدر است؟

دسته‌بند سافت‌مکس (رگرسیون لجستیک چند کلاسی)

- امتیازها. لگاریتم احتمال کلاس‌ها به صورت نرمال نشده!

$$L_i = -\log\left(\frac{e^{s_{y_i}}}{\sum_j e^{s_j}}\right)$$

$$L_i = -\log(0.13) = \mathbf{0.89}$$

	لگاریتم احتمال‌ها به صورت غیرنرمال		احتمال‌ها به صورت غیرنرمال		احتمال‌ها
گربه	3.2	توان رسانی →	24.5	نرمال‌سازی →	0.13
خودرو	5.1		164.0		0.87
قورباغه	-1.7		0.2		0.0

به طور معمول در لحظه شروع، مقادیر W با اعداد تصادفی کوچک مقداردهی می‌شوند و بنابراین تمام امتیازها تقریبا صفر هستند. در این حالت مقدار تابع هزینه کدام است؟

توابع هزینه

$$W = \begin{bmatrix} 0.01 & -0.05 & 0.1 & 0.05 \\ 0.7 & 0.2 & 0.05 & 0.16 \\ 0.0 & -0.45 & -0.2 & 0.03 \end{bmatrix}, \quad x_i = \begin{bmatrix} -15 \\ 22 \\ -44 \\ 56 \end{bmatrix} + b = \begin{bmatrix} 0.0 \\ 0.2 \\ -0.3 \end{bmatrix}$$

$$y_i \quad \boxed{2}$$

hinge loss (SVM)

$$\begin{bmatrix} -2.85 \\ 0.86 \\ 0.28 \end{bmatrix}$$

```
max(0, -2.85 -0.28 + 1) +
max(0,  0.86 -0.28 + 1)
=
1.58
```

cross-entropy loss (Softmax)

$$\begin{bmatrix} -2.85 \\ 0.86 \\ 0.28 \end{bmatrix} \xrightarrow{\text{exp}} \begin{bmatrix} 0.058 \\ 2.36 \\ 1.32 \end{bmatrix} \xrightarrow{\text{normalize}} \begin{bmatrix} 0.016 \\ 0.631 \\ 0.353 \end{bmatrix}$$

```
-log(0.353)
=
1.04
```

مقایسه

- پرسش. فرض کنید یک داده را برداریم و آن را اندکی جابه‌جا کنیم (امتیازهای آن را کمی تغییر دهیم). در مورد هر یک از دو تابع هزینه‌ی داده شده، مقدار هزینه چگونه تغییر می‌کند؟

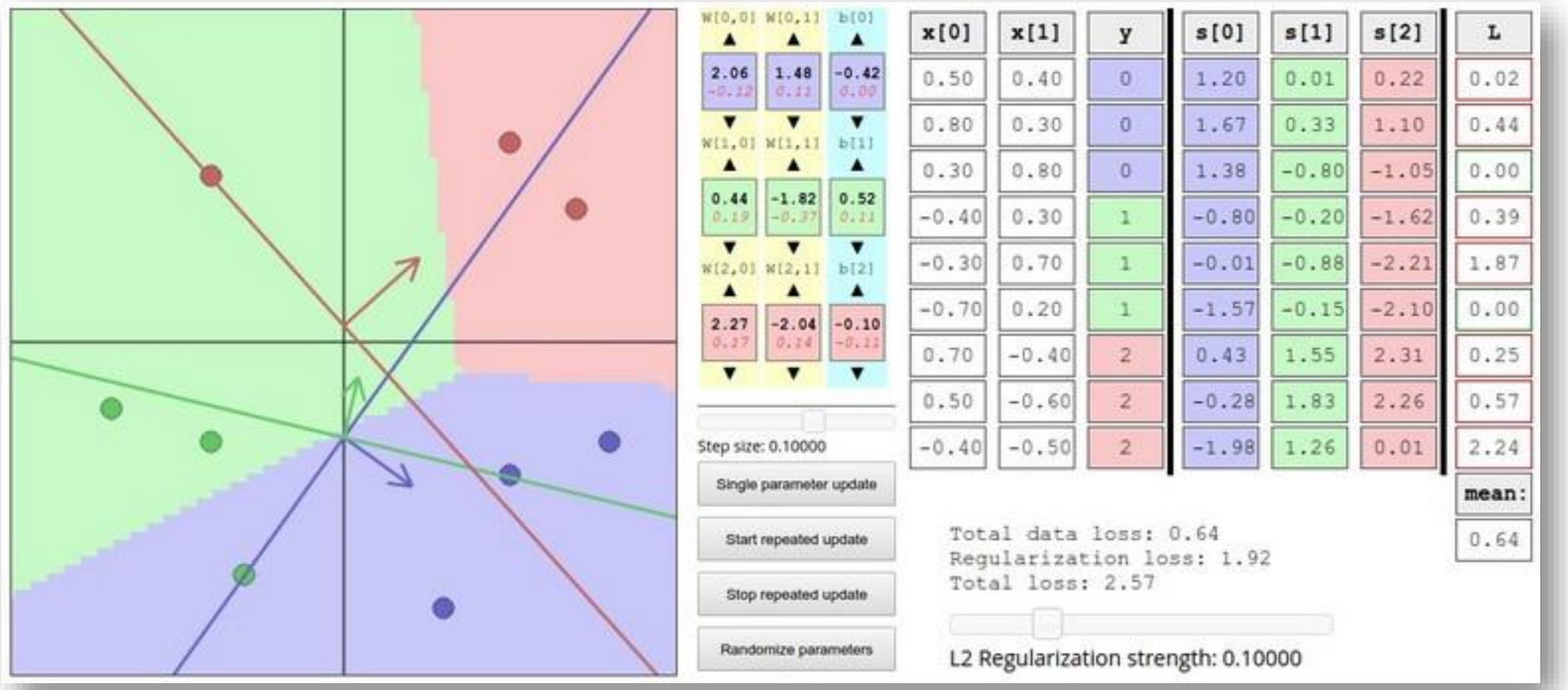
$$L_i = \sum_{j \neq y_i} \max(0, s_j - s_{y_i} + 1)$$

$$L_i = -\log\left(\frac{e^{s_{y_i}}}{\sum_j e^{s_j}}\right)$$

- امتیازها. $[y_i = 0]$

```
[10,   -2,    3]
[10,    9,    9]
[10, -100, -100]
```

اجرای نمایشی



http://vision.stanford.edu/teaching/cs231n/linear-classify-demo/

پیوست: تابع هزینه Cross-Entropy

$$CE(s,y) = -\sum_{i=1}^{C} y_i \log(p_i)$$