1.6

Complete `train` function to perform stochastic gradient descent on the logistic loss function. This will require that you fill in:

1. computing the surrogate loss function at each epoch ($J = \frac{1}{m}\sum J_j$, from the previous part);
2. computing the response ($r^{(j)}$ and gradient associated with each data point $x^{(j)}$, $y^{(j)}$);
3. a stopping criterion consisting of two conditions (stop when either you have reached `stopEpochs` epochs or J has not changed by more than `stopTol` since the last epoch).

Include the complete implementation of `train`.

1.7

Run `train` for your logistic regression classifier on both data sets (A and B). Describe your parameter choices for each dataset (stepsize, etc.) and include plots showing the convergence of the surrogate loss and error rate (e.g., the loss values as a function of epoch during gradient descent), and the final converged classifier with the data (the included `train` function does that for you already).

Problem 2: Shattering

2.1

Consider the data points in Figure 1 which have two real-valued features x_1, x_2. We are also giving a few learners below. For the learners below, $T[z]$ is the sign threshold function

$$T[z] = \begin{cases} +1 & \text{if } z \geq 0 \\ -1 & \text{if } z < 0 \end{cases}$$

The learner parameters $a, b, c, \ldots$ are real-valued scalars, and each data point has two real-valued features x_1, x_2.

Which of the four datasets can be shattered by each learner? Give a brief explanation/justification and use your results to guess the VC dimension of the classifier (you do not have to give a formal proof, just your reasoning)

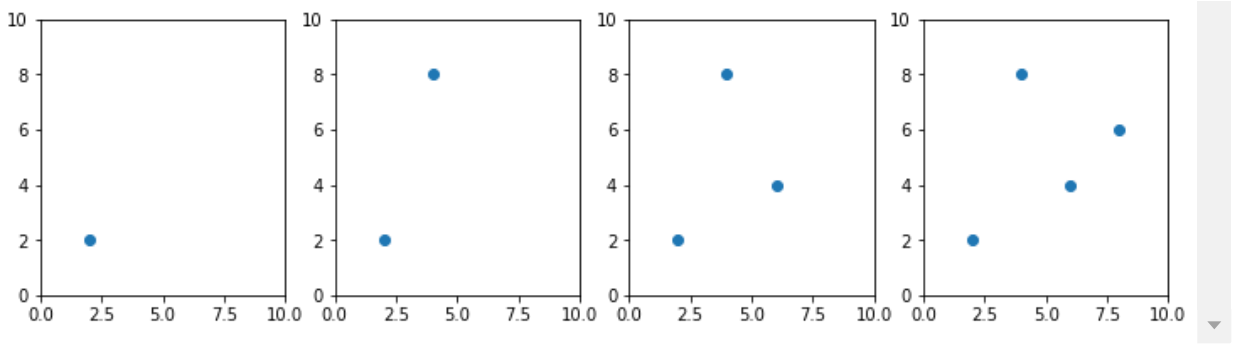
In [38]:
```
fig = plt.figure(figsize=(12, 3))
a = plt.subplot(141);fig.add_axes(a);a.scatter(2,2);a.set_xlim([0,10]);a.set_ylim([0,

b = plt.subplot(142);fig.add_axes(b);b.scatter([2, 4],[2, 8]);b.set_xlim([0,10]);b.se

c = plt.subplot(143);fig.add_axes(c);c.scatter([2, 4, 6],[2, 8, 4]);c.set_xlim([0,10]

d = plt.subplot(144);fig.add_axes(d);d.scatter([2, 4, 6, 8],[2, 8, 4, 6]);d.set_xlim(
```

Out[38]:
```
(0.0, 10.0)
```



2.1.1

$T(a + bx_1)$

This can shatter plots A and B.

A

- The learner is able to correctly categorize point (2, 2) when:

$(2, 2)$	a value	b value
$+$	0	0
-	-1	-1

B

- The learner is able to correctly categorize the two points when:

$(2, 2)$	$(4, 8)$	a value	b value
$+$	$+$	0	0
$+$	-	-1	-1
-	$+$	2	-1
-	-	-3	1

It cannot shatter set C: If (2, 2) is positive, (4, 8) is negative, and (6, 4) is positive, there is no way to multiply and add numbers that would result in 2 and 6 both being larger than 4.

I think that the VC dimension of this classifier is 2.

2.1.2

$T((a * b)x_1 + (\frac{c}{a})x_2)$

A

- The learner is able to correctly categorize point (2, 2) when:

$(2, 2)$	a value	b value	c value
$+$	1	0	0
-	-1	1	1

B

- The learner is able to correctly categorize points (2, 2) and (4, 8):

$(2, 2)$	$(4, 8)$	a value	b value	c value
$+$	$+$	1	0	0
$+$	-	1	1	-1
-	$+$	1	-1	$\frac{1}{2}$
-	-	1	1	-2

2.1.3

$T((x_1 - a)^2 + (x_2 - b)^2 + c)$

This can shatter plots A and B.

A

- The learner is able to correctly categorize point (2, 2) when:

$(2, 2)$	a value	b value	c value
$+$	2	2	1
-	2	2	-1

B

- The learner is able to correctly categorize points (2, 2) and (4, 8) when:

$(2, 2)$	$(4, 8)$	a value	b value	c value
$+$	$+$	0	0	0
$+$	-	4	8	-1
-	$+$	2	2	-1
-	-	3	5	-11

The learner cannot shatter plot C because it fails when (2, 2) and (4, 8) are negative and (6, 4) is positive.

I think the VC Dimension of the classifier is 2.

2.1.4

$T(a + bx_1 + cx_2) \times T(d + bx_1 + cx_2)$

Hint: The two equations are two parallel lines

This can shatter plots A, B

A

- The learner is able to correctly categorize point (2, 2) when:

$(2, 2)$	a value	b value	c value	d value
$+$	0	1	1	0
-	-1	0	0	-1

B

- The learner is able to correctly categorize points (2, 2) and (4, 8) when:

$(2, 2)$	$(4, 8)$	a value	b value	c value	d value
$+$	$+$	0	0	0	0
$+$	-	-1	2	-1	-1
-	$+$	-1	-1	1	-1
-	-	0	-1	-1	0

C

- The learner is able to correctly categorize points (2, 2), (4, 8), and (6, 4) when:

$(2, 2)$	$(4, 8)$	$(4, 8)$	a value	b value	c value	d value
$+$	$+$	$+$	0	1	1	0
$+$	$+$	-	0	-1	1	0
$+$	-	$+$	0	1	-1	0
$+$	-	-	2	1	-2	2
-	$+$	$+$	-2	1	0	-2
-	$+$	-	-1	-1	1	-1
-	-	$+$	-4	1	0	-4
-	-	-	0	-1	-1	0

The learner cannot shatter D because it cannot correctly categorize the data when (2, 2), (4, 8), and (8, 6) are positive and (6, 4) is negative.

I think the VC dimension of this classifier is 3.

Statement of Collaboration