Here we provide the pseudo code for linear SVM. We assume that the following parameters are known.

- $\mathbf{w}, w_0$: the weights/parameters to be updated by SGD
- m: the size of minibatch
- n: the size of data
- l: the learnig rate

And the gradient under different cases are as follows: (you need to compute $\mathbf{f}_1, \mathbf{f}_2, g_1, g_2$ by taking the derivatives of the objective function)

- $\frac{\partial f}{\partial \mathbf{w}} = \begin{cases} \mathbf{f}_1, if\ 1 - y^t(\mathbf{w}^T x + w_0) \leq 0 \\ \mathbf{f}_2, if\ 1 - y^t(\mathbf{w}^T x + w_0) > 0 \end{cases}$
- $\frac{\partial f}{\partial w_0} = \begin{cases} g_1, if\ 1 - y^t(\mathbf{w}^T x + w_0) \leq 0 \\ g_2, if\ 1 - y^t(\mathbf{w}^T x + w_0) > 0 \end{cases}$

In SVM, the Labels y are 1/-1 rather than 0/1, you need to make corresponding adjustment with the input data

Algorithm 1 fit(X, y)

for $epoch = 1, 2, \ldots$ **do**
  for $batch = 1, 2, \ldots, \frac{n}{m}$ **do**
    Compute the gradient ∂f (you need to choose between $\mathbf{f}_1, \mathbf{f}_2, g_1, g_2$ by considering $1 - y^t(\mathbf{w}^T x + w_0)$) (notice the x used here is the set that containing only the data in the corresponding mini-batch.
    Update weight: $\mathbf{w} \leftarrow \mathbf{w} - l\frac{\partial f}{\partial \mathbf{w}}, w_0 \leftarrow w_0 - l\frac{\partial f}{\partial w_0}$
  end for
  (Optional) Early Stopping to terminate training when the models has already converge before the end of pre-defined number of epochs.
end for

Algorithm 2 predict(X)

for $x = 1, 2, \ldots, n$ **do** (we iterate over all the testing data instances)
  score=$\mathbf{w^t}x + w_0$
  if score>0 **then**
    Set corresponding label y as 1
  else
    Set corresponding label y as -1
  end if
end for