```
clear
parallel.gpu.enableCUDAForwardCompatibility(true)
canUseGPU
```

```
ans = logical
   1
```

```
gpuDevice
```

```
ans =
  CUDADevice with properties:

                Name: 'NVIDIA RTX PRO 2000 Blackwell Generation Laptop GPU'
               Index: 1 (of 1)
    ComputeCapability: '12.0'
          DriverModel: 'WDDM'
          TotalMemory: 8546549760 (8.55 GB)
      AvailableMemory: 6905880576 (6.91 GB)
      DeviceAvailable: true
       DeviceSelected: true

  Show all properties.
```

```
run("winedata.m")
```

```
reviewText = winemag_data_first150k.description;
documents = tokenizedDocument(reviewText);
documents = lower(documents);
documents = removeStopWords(documents);
bag = bagOfWords(documents)
```

```
bag =
  bagOfWords with properties:

        NumWords: 41427
          Counts: [150930×41427 double]
      Vocabulary: ["tremendous"    "100"    "%"    "varietal"    "wine"    "hails"    "oakville"    "aged"    "three
    NumDocuments: 150930
```

```
X = tfidf(bag);
```

```
size(X)
```

```
ans = 1×2
      150930       41427
```

```
nnz(X)
```

```
ans =
3961510
```

```
elements = 50;
[U, S, V] = svds(X, elements);
Z = U*S;
```

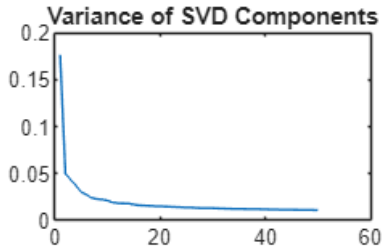
```
vocab = bag.Vocabulary;
pc1 = V(:,1);
pc2 = V(:,2);

[~, idxPos] = sort(pc1, 'descend');
topPosWords_pc1 = vocab(idxPos(1:15))
```

```
topPosWords_pc1 = 1×15 string
"%"          "fruit"      "wine"       "black"      "tannins"    "finish"     "a ···
```

```
[~, idxNeg] = sort(pc1, 'ascend');
topNegWords_pc1 = vocab(idxNeg(1:15))
```

```
topNegWords_pc1 = 1×15 string
"judi"       "flom"       "sorensen"   "gluey-sweet""mawkish"    "strawberry-ba… ···
```

```
figure
singVals = diag(S).^2;
explained = singVals / sum(singVals);
plot(explained(1:50));
title("Variance of SVD Components");
```

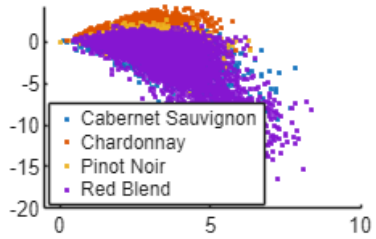



```
labels = categorical(winemag_data_first150k.variety);
[vals,~,idx] = unique(labels);
counts = histcounts(idx, 1:numel(vals)+1);
[~,order] = sort(counts, 'descend');
topN = 4;
keepVarieties = vals(order(1:topN))
```

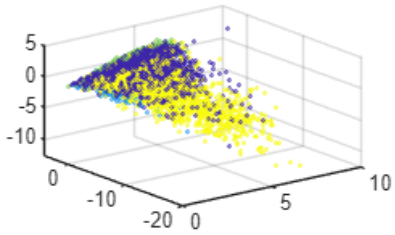
```
keepVarieties = 4×1 categorical
Chardonnay
Pinot Noir
Cabernet Sauvignon
Red Blend
```

```
mask = ismember(labels, keepVarieties);
Zsmall = Z(mask, :);
labelSmall = removecats(labels(mask));
```

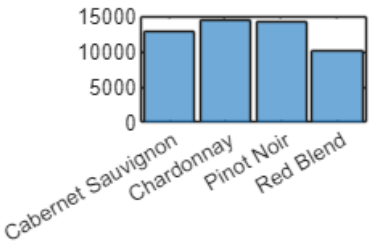
```
figure
gscatter(Zsmall(:,1), Zsmall(:,2), labelSmall);
```




```
figure
scatter3(Zsmall(:,1),Zsmall(:,2),Zsmall(:,3),1,labelSmall)
```



```
features = Z(:,1:elements);

histogram(labelSmall)
```




```
textSmall  = winemag_data_first150k.description(mask);
labelSmall = removecats(labels(mask));

cv = cvpartition(labelSmall,'HoldOut',0.2);

XtrainText = string(textSmall(training(cv)));
Ytrain     = labelSmall(training(cv));

XtestText  = string(textSmall(test(cv)));
Ytest      = labelSmall(test(cv));
```

```
%mdl = fitcecoc(Xtrain, Ytrain);  % multiclass SVM
%Ypred = predict(mdl, Xtest);

%accuracy_svm = mean(Ypred == Ytest);
%disp(accuracy_svm);
```

```
unique(string(labelSmall))
```

```
ans = 4×1 string
"Cabernet Sauvi…
"Chardonnay"
"Pinot Noir"
"Red Blend"
```

```
mdl = bertDocumentClassifier(ClassNames=unique(string(labelSmall)))
```

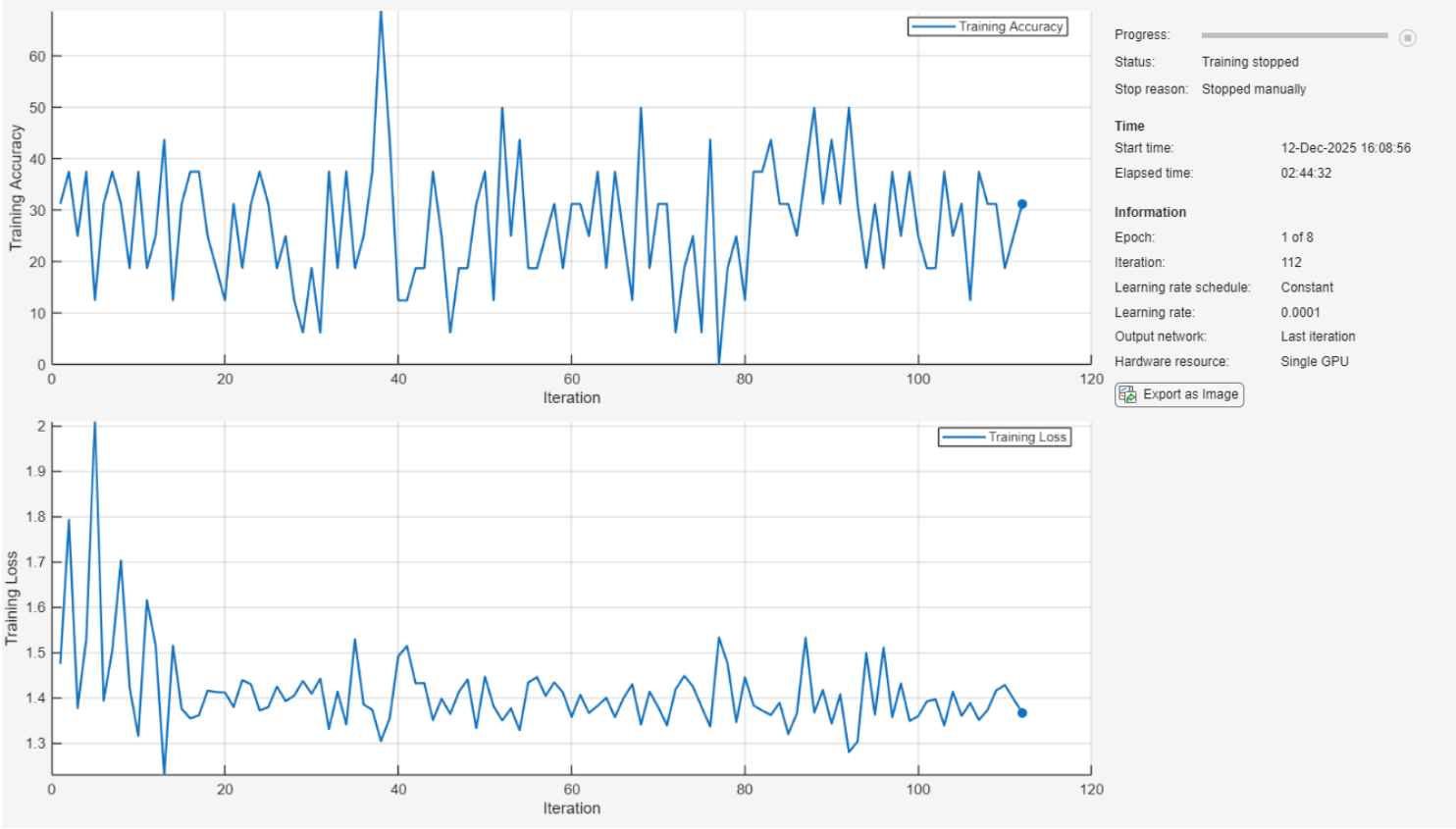
```
mdl =
  bertDocumentClassifier with properties:

       Network: [1×1 dlnetwork]
     Tokenizer: [1×1 bertTokenizer]
    ClassNames: ["Cabernet Sauvignon"    "Chardonnay"    "Pinot Noir"    "Red Blend"]
```

```
options = trainingOptions("adam", ...
    ExecutionEnvironment="gpu", ...
    MaxEpochs=8, ...
    MiniBatchSize=16, ...
    InitialLearnRate=1e-4, ...
    Shuffle="every-epoch", ...
    Plots="training-progress", ...
    Metrics="accuracy", ...
    Verbose=true);
Xtext = reviewText(training(cv))
```

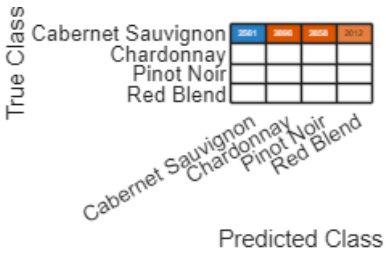
```
Xtext = 41308×1 string
"This tremendous 100% varietal wine hails from Oakville and was aged over three  ···
"Ripe aromas of fig, blackberry and cassis are softened and sweetened by a slath
"Deep, dense and pure from the opening bell, this Toro is a winner. Aromas of da
"Lush cedary black-fruit aromas are luxe and offer notes of marzipan and vanilla
"The producer sources from two blocks of the vineyard for this wine—one at a hig
"Elegance, complexity and structure come together in this drop-dead gorgeous win
"From 18-year-old vines, this supple well-balanced effort blends flavors of moch
"A standout even in this terrific lineup of 2015 releases from Patricia Green, t
"This wine is in peak condition. The tannins and the secondary flavors dominate
"With its sophisticated mix of mineral, acid and tart fruits, this seductive eff
     ⋮
```

```
mdl = trainBERTDocumentClassifier(string(Xtext), string(Ytrain),mdl,options);
```

Iteration	Epoch	TimeElapsed	LearnRate	TrainingLoss	TrainingAccuracy
1	1	00:00:01	0.0001	2.0216	18.75




```
YTest =labelSmall(test(cv));
TTest = classify(mdl,reviewText(test(cv)));

figure
confusionchart(TTest,YTest)
```




```
accuracy = mean(TTest == YTest)
```

```
accuracy =
0.2480
```