Project 1 - Part B

Code

Hide

```
load("../data/features.RData")
head(features_df)
```

	Ind...	Panel	
	<int>	<chr>	
...data.ProjectData_ForClass_v2.3.s1.img.s1.jpg	1	scroll_001_panel_01	0.(
...data.ProjectData_ForClass_v2.3.s1.img.s1.jpg.1	2	scroll_001_panel_02	0.(
...data.ProjectData_ForClass_v2.3.s1.img.s1.jpg.2	3	scroll_001_panel_03	0.(
...data.ProjectData_ForClass_v2.3.s1.img.s2.jpg	4	scroll_002_panel_01	0.5
...data.ProjectData_ForClass_v2.3.s1.img.s3.jpg	5	scroll_003_panel_01	0.5
...data.ProjectData_ForClass_v2.3.s1.img.s3.jpg.1	6	scroll_003_panel_02	0.5

6 rows | 1-5 of 5 columns

Hide

```
annotation <- read.csv("../data/Annotation.csv", stringsAsFactors = FALSE)


annotation$Tree <- as.numeric(annotation$Tree)
annotation$Animal <- as.numeric(annotation$Animal)
annotation$Mythological <- as.numeric(annotation$Mythological)


merged_df <- merge(features_df, annotation, by.x = "Panel", by.y = "Panel")


head(merged_df)
```

	Panel	Index.x	R_mean	G_mean	B_mean	Index.y	Tr...	Ani...
	<chr>	<int>	<dbl>	<dbl>	<dbl>	<int>	<dbl>	<dbl>
1	scroll_001_panel_01	1	0.6009400	0.5243191	0.4382841	1	1	1
2	scroll_001_panel_02	2	0.6009400	0.5243191	0.4382841	2	1	1

3	scroll_001_panel_03	3	0.6009400	0.5243191	0.4382841	3	1	1
4	scroll_002_panel_01	4	0.5553909	0.4939770	0.4114456	4	*NA*	*NA*
5	scroll_003_panel_01	5	0.5792157	0.5425198	0.4848491	5	1	1
6	scroll_003_panel_02	6	0.5792157	0.5425198	0.4848491	6	1	1

6 rows | 1-9 of 11 columns

Hide

```
annotation <- read.csv("../data/Annotation.csv", stringsAsFactors = FALSE)


labels_df <- annotation[, c("Index", "Panel", "Tree", "Animal", "Mythological.Charact
er")]
colnames(labels_df) <- c("Index", "Panel", "Tree_Label", "Animal_Label", "Mythology_L
abel")


labels_df <- na.omit(labels_df)
head(labels_df)
```

	Index	Panel	Tree_Label	Animal_Label	Mythology_Label
	<int>	<chr>	<int>	<int>	<int>
1	1	scroll_001_panel_01	1	1	1
2	2	scroll_001_panel_02	1	1	1
3	3	scroll_001_panel_03	1	1	1
5	5	scroll_003_panel_01	1	1	1
6	6	scroll_003_panel_02	1	1	1
8	8	scroll_005_panel_01	1	1	0

6 rows

Hide

```
merged_df <- merge(features_df, labels_df, by = c("Index", "Panel"))
save(merged_df, file = "../data/merged_features_labels.RData")
```

Hide

```
load("../data/merged_features_labels.RData")


str(merged_df)
```

```
'data.frame':   28 obs. of  8 variables:
 $ Index          : int  1 10 100 101 102 103 11 12 16 17 ...
 $ Panel          : chr  "scroll_001_panel_01" "scroll_005_panel_03" "scroll_017_pane
l_03" "scroll_017_panel_04" ...
 $ R_mean         : num  0.601 0.557 0.506 0.506 0.506 ...
 $ G_mean         : num  0.524 0.471 0.457 0.457 0.457 ...
 $ B_mean         : num  0.438 0.391 0.379 0.379 0.379 ...
 $ Tree_Label     : int  1 1 1 1 0 0 1 1 1 0 ...
 $ Animal_Label   : int  1 1 0 0 0 0 1 1 1 1 ...
 $ Mythology_Label: int  1 0 1 1 1 1 0 0 0 1 ...
```

Hide

```
tree_data <- merged_df[, c("R_mean", "G_mean", "B_mean")]
tree_labels <- merged_df$Tree_Label

animal_data <- merged_df[, c("R_mean", "G_mean", "B_mean")]
animal_labels <- merged_df$Animal_Label

myth_data <- merged_df[, c("R_mean", "G_mean", "B_mean")]
myth_labels <- merged_df$Mythology_Label


save(tree_data, tree_labels, animal_data, animal_labels, myth_data, myth_labels,
     file = "../data/datasets.RData")

# Preview
table(tree_labels)
```

```
tree_labels
 0  1
11 17
```

Hide

```
table(animal_labels)
```

```
animal_labels
 0  1
14 14
```

Hide

```
table(myth_labels)
```

```
myth_labels
 0  1
 7 21
```

Hide

```
train_sgd <- function(X, y, loss_type = "logistic", lr = 0.01, epochs = 100) {
  n <- nrow(X)
  p <- ncol(X)
  theta <- rep(0, p)
  loss_history <- numeric(epochs)

  for (epoch in 1:epochs) {
    for (i in 1:n) {
      xi <- X[i, , drop = FALSE]
      yi <- y[i]
      score <- as.numeric(xi %*% theta)

      if (loss_type == "logistic") {
        grad <- as.numeric(-yi * xi / (1 + exp(yi * score)))
      } else if (loss_type == "cross-entropy") {
        prob <- 1 / (1 + exp(-score))
        grad <- as.numeric((prob - yi) * xi)
      }

      theta <- theta - lr * grad
    }


    probs <- 1 / (1 + exp(-X %*% theta))
    if (loss_type == "logistic") {
      loss_history[epoch] <- mean(log(1 + exp(-y * (X %*% theta))))
    } else if (loss_type == "cross-entropy") {
      eps <- 1e-6
      loss_history[epoch] <- -mean(y * log(probs + eps) + (1 - y) * log(1 - probs + e
ps))
    }
  }

  list(theta = theta, loss_history = loss_history)
}
```

Hide

```
X_tree <- as.matrix(tree_data)
y_tree <- as.numeric(tree_labels)

X_animal <- as.matrix(animal_data)
y_animal <- as.numeric(animal_labels)

X_myth <- as.matrix(myth_data)
y_myth <- as.numeric(myth_labels)
```

Hide

```
X_tree <- as.matrix(tree_data)
y_tree <- as.numeric(tree_labels)

X_animal <- as.matrix(animal_data)
y_animal <- as.numeric(animal_labels)

X_myth <- as.matrix(myth_data)
y_myth <- as.numeric(myth_labels)
```

Hide

```
dim(X_tree)
```

```
[1] 28  3
```

Hide

```
length(y_tree)
```

```
[1] 28
```

Hide

```
sgd_tree_logistic <- train_sgd(X_tree, y_tree, loss_type = "logistic")
sgd_tree_entropy  <- train_sgd(X_tree, y_tree, loss_type = "cross-entropy")

sgd_animal_logistic <- train_sgd(X_animal, y_animal, loss_type = "logistic")
sgd_animal_entropy  <- train_sgd(X_animal, y_animal, loss_type = "cross-entropy")

sgd_myth_logistic <- train_sgd(X_myth, y_myth, loss_type = "logistic")
sgd_myth_entropy  <- train_sgd(X_myth, y_myth, loss_type = "cross-entropy")
```

Hide

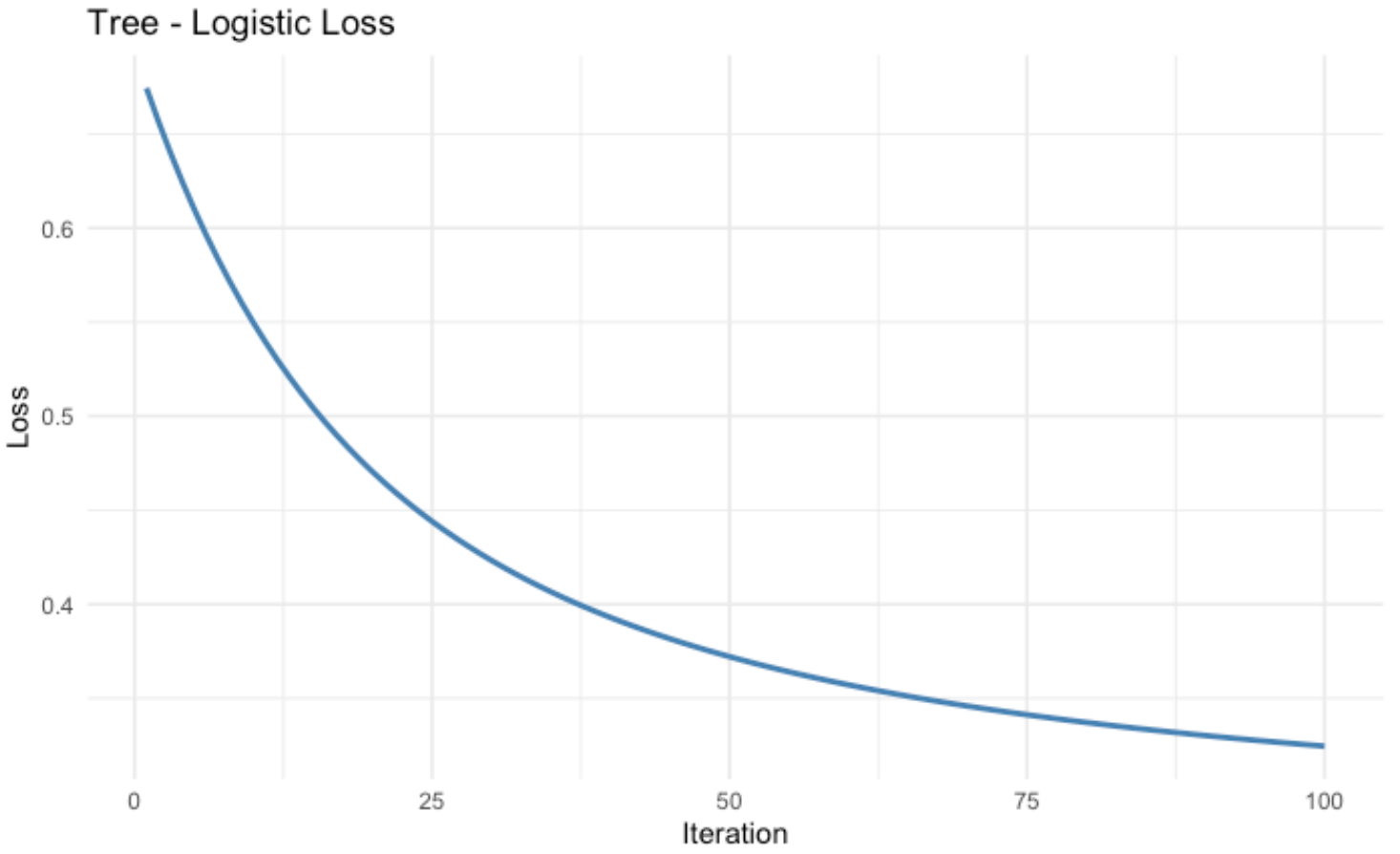
```
library(ggplot2)

plot_loss_curve <- function(model, title) {
  loss_df <- data.frame(
    Iteration = seq_along(model$loss_history),
    Loss = model$loss_history
  )

  ggplot(loss_df, aes(x = Iteration, y = Loss)) +
    geom_line(color = "steelblue", linewidth = 1) +
    ggtitle(title) +
    theme_minimal()
}
```

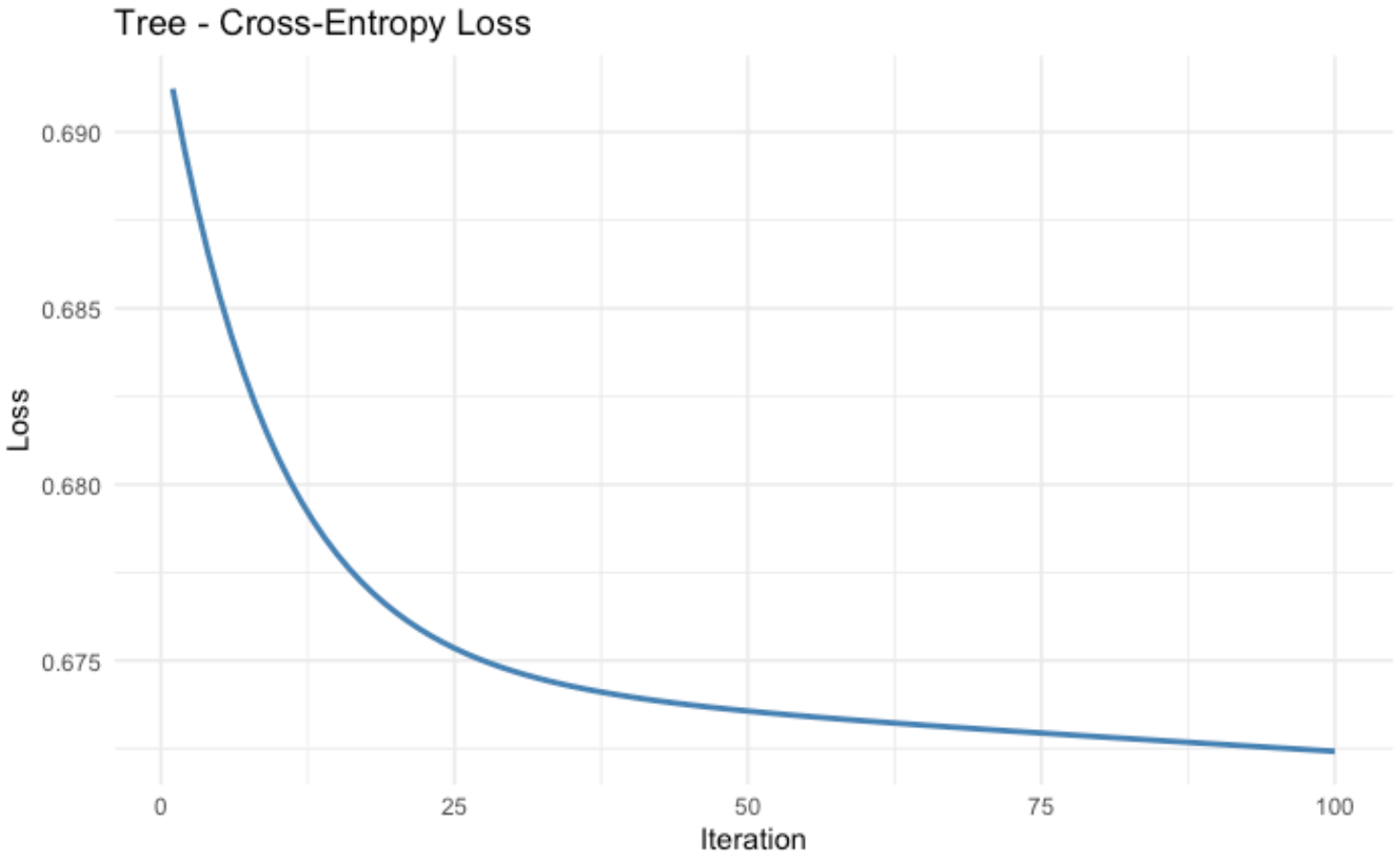
Hide

```
plot_loss_curve(sgd_tree_logistic, "Tree - Logistic Loss")
```



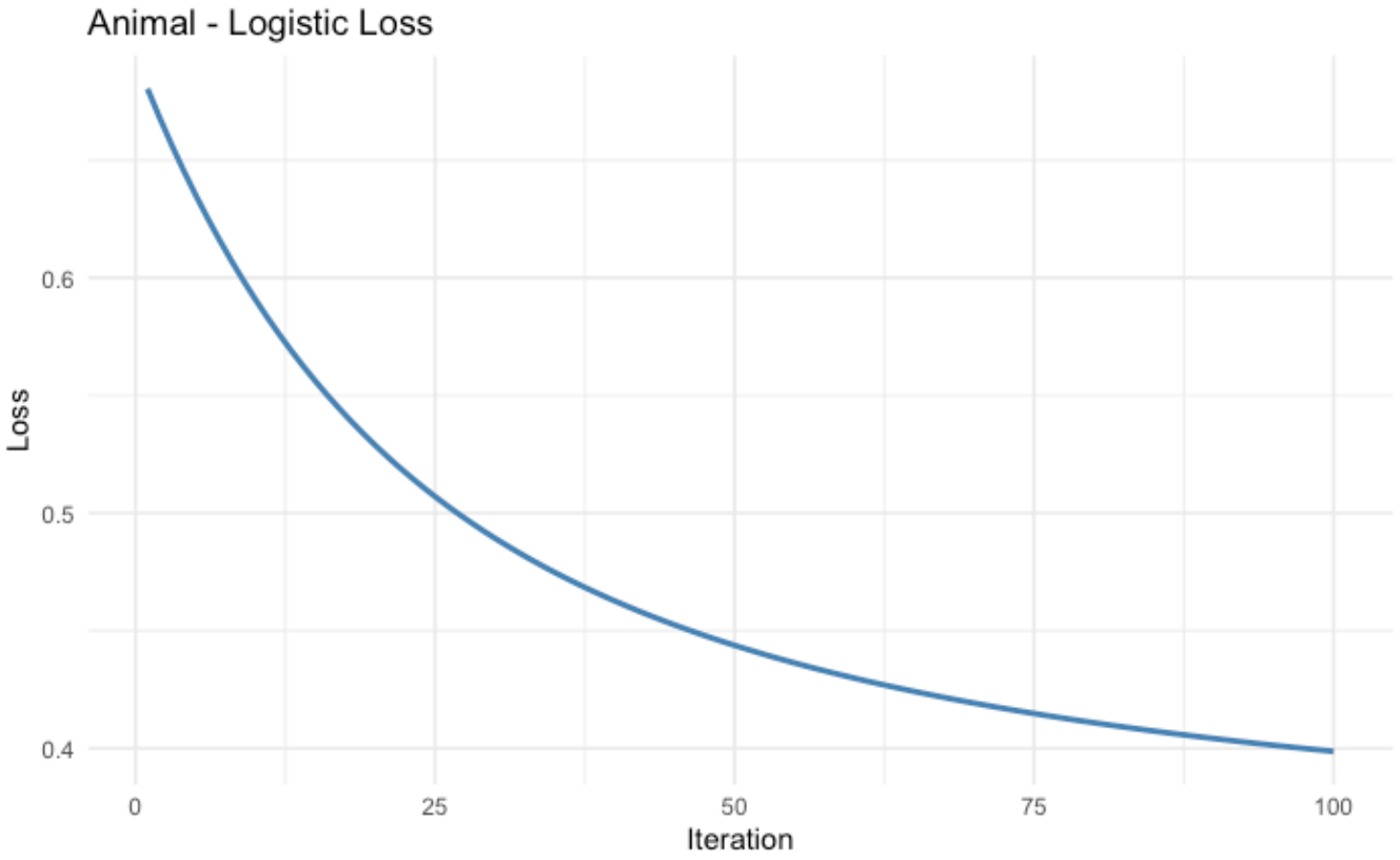
Hide

```
plot_loss_curve(sgd_tree_entropy, "Tree - Cross-Entropy Loss")
```



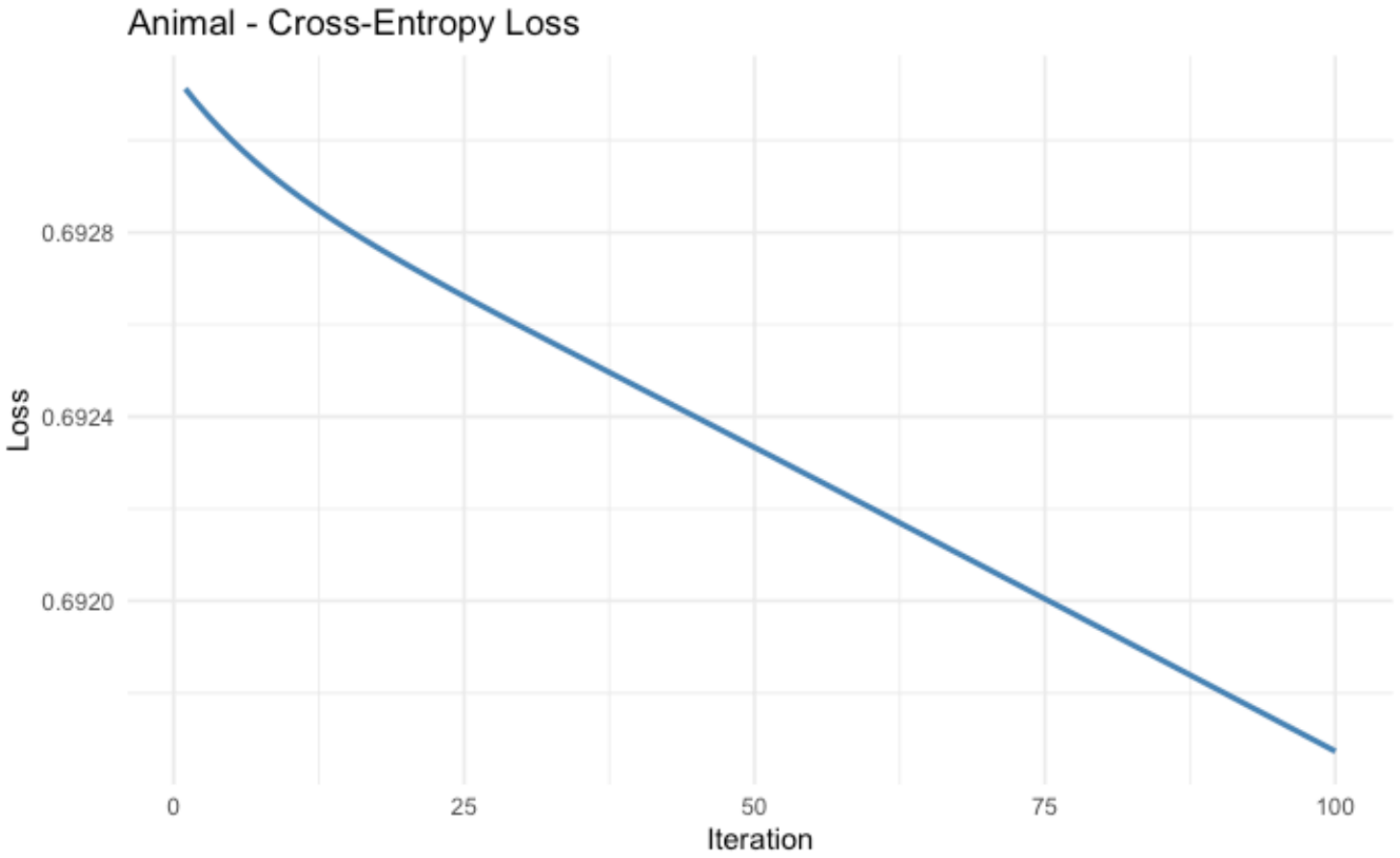
Hide

```
plot_loss_curve(sgd_animal_logistic, "Animal - Logistic Loss")
```



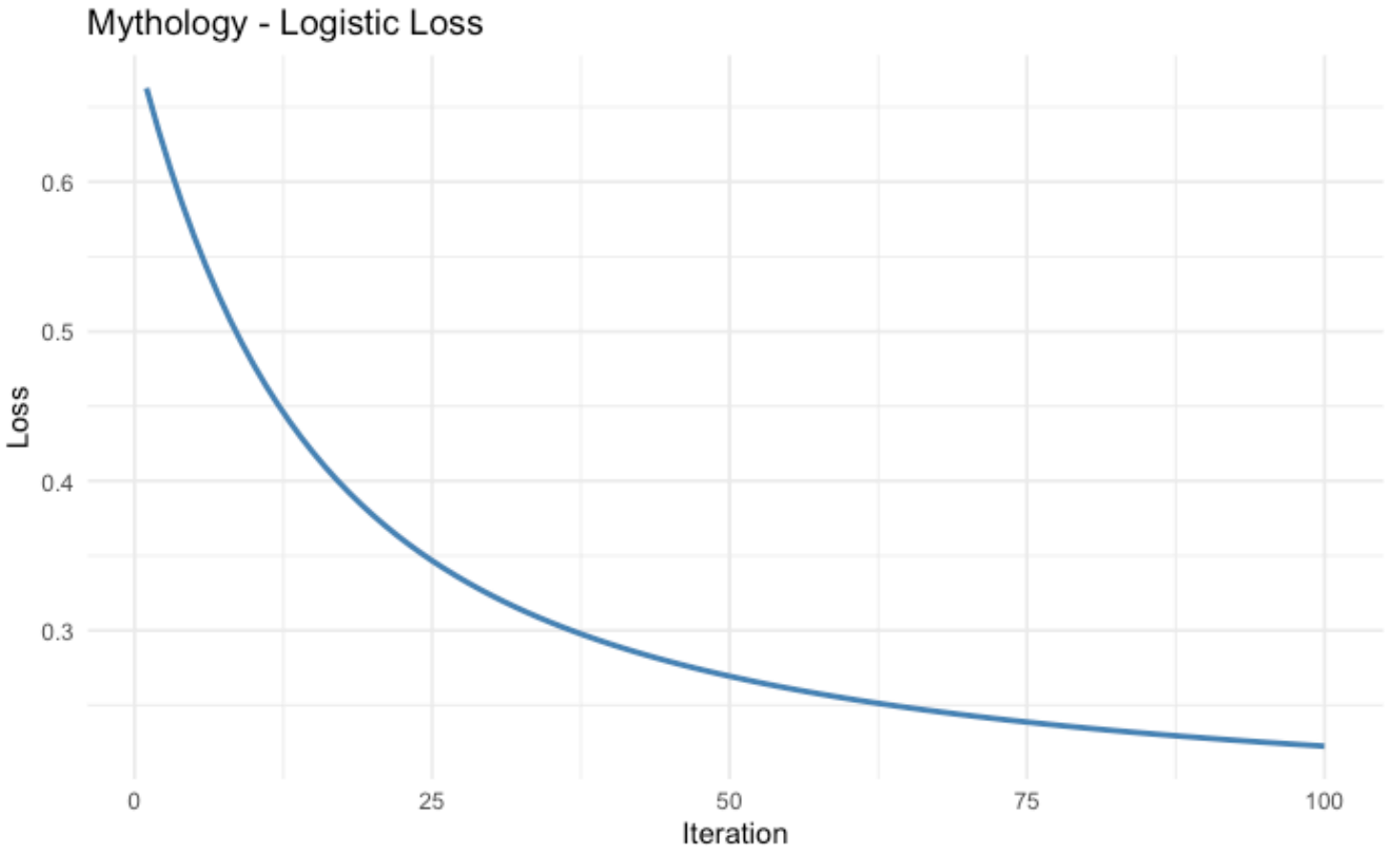
Hide

```
plot_loss_curve(sgd_animal_entropy, "Animal - Cross-Entropy Loss")
```



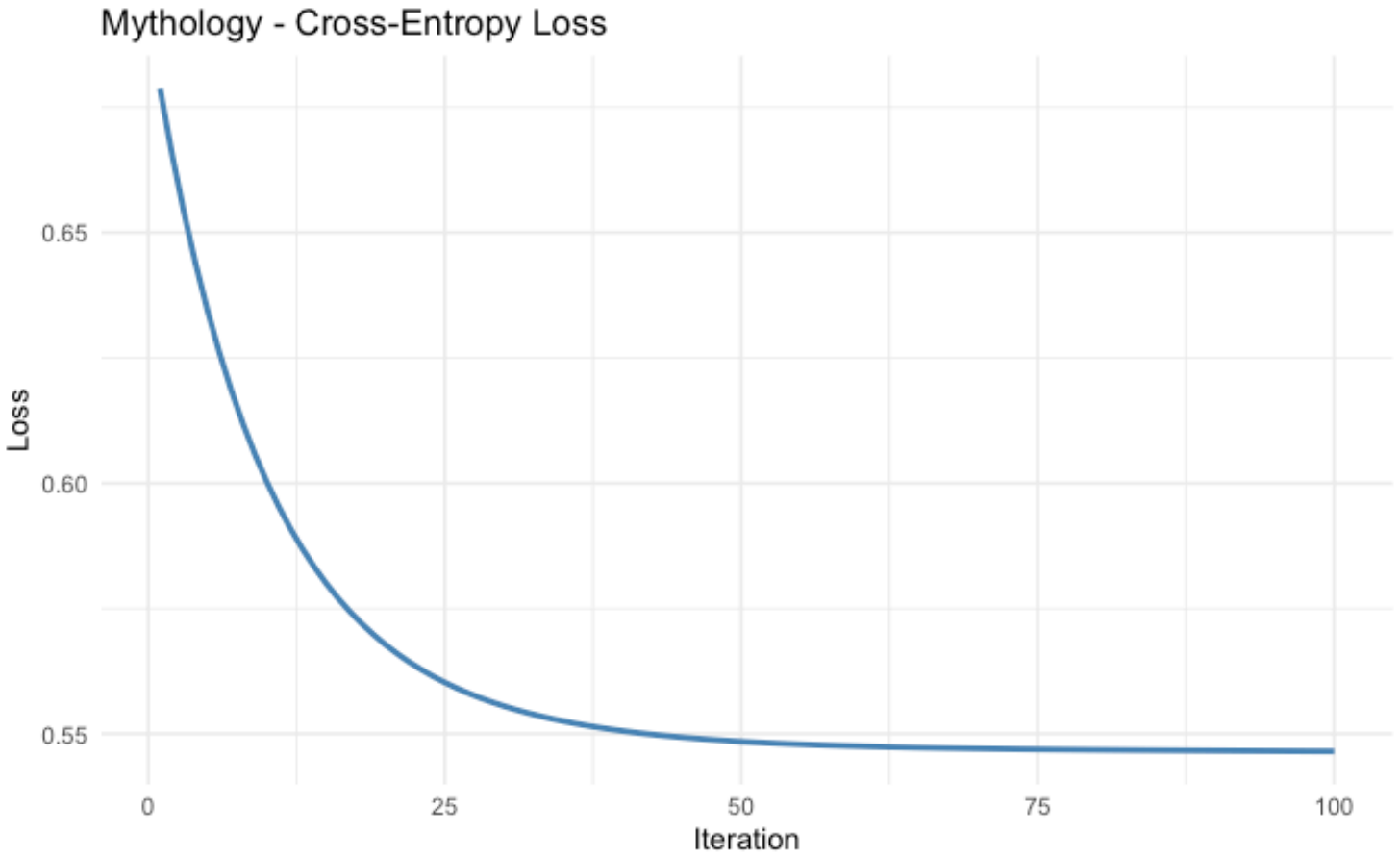
Hide

```
plot_loss_curve(sgd_myth_logistic, "Mythology - Logistic Loss")
```



Hide

```
plot_loss_curve(sgd_myth_entropy, "Mythology - Cross-Entropy Loss")
```



Hide

```
predict_labels <- function(model, X) {
  probs <- 1 / (1 + exp(-X %*% model$theta))
  preds <- ifelse(probs >= 0.5, 1, 0)
  return(preds)
}
```

Hide

```
compute_accuracy <- function(y_true, y_pred) {
  mean(y_true == y_pred)
}
```

Hide

```
# Tree
tree_preds_log <- predict_labels(sgd_tree_logistic, X_tree)
tree_preds_entropy <- predict_labels(sgd_tree_entropy, X_tree)

cat("Tree Accuracy (Logistic):", compute_accuracy(y_tree, tree_preds_log), "\n")
```

```
Tree Accuracy (Logistic): 0.6071429
```

Hide

```
cat("Tree Accuracy (Cross-Entropy):", compute_accuracy(y_tree, tree_preds_entropy), "
\n")
```

```
Tree Accuracy (Cross-Entropy): 0.6071429
```

Hide

```
# Animal
animal_preds_log <- predict_labels(sgd_animal_logistic, X_animal)
animal_preds_entropy <- predict_labels(sgd_animal_entropy, X_animal)

cat("Animal Accuracy (Logistic):", compute_accuracy(y_animal, animal_preds_log), "\
n")
```

```
Animal Accuracy (Logistic): 0.5
```

Hide

```
cat("Animal Accuracy (Cross-Entropy):", compute_accuracy(y_animal, animal_preds_entro
py), "\n")
```

```
Animal Accuracy (Cross-Entropy): 0.5
```

Hide

```
# Mythology
myth_preds_log <- predict_labels(sgd_myth_logistic, X_myth)
myth_preds_entropy <- predict_labels(sgd_myth_entropy, X_myth)

cat("Mythology Accuracy (Logistic):", compute_accuracy(y_myth, myth_preds_log), "\n")
```

```
Mythology Accuracy (Logistic): 0.75
```

Hide

```
cat("Mythology Accuracy (Cross-Entropy):", compute_accuracy(y_myth, myth_preds_entrop
y), "\n")
```

```
Mythology Accuracy (Cross-Entropy): 0.75
```

Hide

```
save(sgd_tree_logistic, sgd_tree_entropy,
     sgd_animal_logistic, sgd_animal_entropy,
     sgd_myth_logistic, sgd_myth_entropy,
     file = "../data/sgd_models.RData")
```