In [74]:
```python
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from lazypredict.Supervised import LazyClassifier
from sklearn.model_selection import cross_val_score
from sklearn.neighbors import KNeighborsClassifier
import matplotlib.pyplot as plt
from sklearn.model_selection import learning_curve
from sklearn.metrics import roc_curve, auc
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import classification_report, confusion_matrix, accuracy_score
import warnings

# Load
df_cancer = pd.read_csv("C:/Users/ali/Desktop/osdabig/questionno3/datasetosda/hh/bcancer3.csv")
df_cancer.rename(columns={"target": "diagnosis"}, inplace=True)

# Separate features and target variable
X = df_cancer.drop('diagnosis', axis=1)
y = df_cancer['diagnosis']

# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=55)

# LazyClassifier
clf = LazyClassifier(verbose=0, ignore_warnings=True, custom_metric=None)
models, predictions = clf.fit(X_train, X_test, y_train, y_test)
models
```

```
 86%|██████████████████████████████████████████████████████████████████████           | 25/29 [00:05<00:00,  4.
19it/s]
```

```
[LightGBM] [Info] Number of positive: 154, number of negative: 245
[LightGBM] [Info] Auto-choosing col-wise multi-threading, the overhead of testing was 0.000422 seconds.
You can set `force_col_wise=true` to remove the overhead.
[LightGBM] [Info] Total Bins 3973
[LightGBM] [Info] Number of data points in the train set: 399, number of used features: 30
[LightGBM] [Info] [binary:BoostFromScore]: pavg=0.385965 -> initscore=-0.464306
[LightGBM] [Info] Start training from score -0.464306
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

```
100%|██████████████████████████████████████████████████████████████████████████| 29/29 [00:06<00:00,  4.
73it/s]
```

```
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
[LightGBM] [Warning] No further splits with positive gain, best gain: -inf
```

Out[74]:

Model	Accuracy	Balanced Accuracy	ROC AUC	F1 Score	Time Taken
LogisticRegression	1.00	1.00	None	1.00	0.11
SVC	0.99	0.99	None	0.99	0.06
LabelSpreading	0.98	0.98	None	0.98	0.09
LabelPropagation	0.98	0.98	None	0.98	0.09
LinearSVC	0.98	0.98	None	0.98	0.06
Perceptron	0.98	0.97	None	0.98	0.05
SGDClassifier	0.97	0.97	None	0.97	0.05
KNeighborsClassifier	0.96	0.96	None	0.96	0.07
LGBMClassifier	0.96	0.96	None	0.96	0.33
ExtraTreesClassifier	0.96	0.96	None	0.96	0.49
CalibratedClassifierCV	0.96	0.95	None	0.96	0.17
AdaBoostClassifier	0.94	0.94	None	0.94	1.57
QuadraticDiscriminantAnalysis	0.94	0.94	None	0.94	0.22
BernoulliNB	0.93	0.93	None	0.93	0.05
BaggingClassifier	0.93	0.93	None	0.93	0.50
RandomForestClassifier	0.93	0.93	None	0.93	1.53
PassiveAggressiveClassifier	0.92	0.92	None	0.92	0.05
NuSVC	0.92	0.92	None	0.92	0.11
RidgeClassifier	0.93	0.92	None	0.93	0.05
RidgeClassifierCV	0.93	0.92	None	0.93	0.07
GaussianNB	0.91	0.91	None	0.91	0.05
NearestCentroid	0.92	0.91	None	0.92	0.06
DecisionTreeClassifier	0.91	0.91	None	0.91	0.09
LinearDiscriminantAnalysis	0.92	0.90	None	0.92	0.06
ExtraTreeClassifier	0.90	0.90	None	0.90	0.04
DummyClassifier	0.60	0.50	None	0.45	0.04

In [75]:

```python
from sklearn.model_selection import train_test_split
from sklearn.neighbors import KNeighborsClassifier
from sklearn.model_selection import cross_val_score
from sklearn.metrics import confusion_matrix, classification_report
import numpy as np

def tune_knn_parameters(X, y):
    # Split the data into training and testing sets
    X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=40)
```

```
    # Test different values of k
    k_values = range(1, 20)
    for k in k_values:
        knn = KNeighborsClassifier(n_neighbors=k)
        scores = cross_val_score(knn, X_train, y_train, cv=5)
        print(f'k={k}, Mean Accuracy: {np.mean(scores)}')

def lazy_classification(X_train, y_train, X_test, y_test, k_value):
    # Train k-NN with the selected k
    knn = KNeighborsClassifier(n_neighbors=k_value)
    knn.fit(X_train, y_train)

    # Predictions on the test set
    y_pred = knn.predict(X_test)

    # Confusion Matrix
    cm = confusion_matrix(y_test, y_pred)
    print("Confusion Matrix:")
    print(cm)

    # Classification Report
    cr = classification_report(y_test, y_pred)
    print("\nClassification Report:")
    print(cr)

    # Evaluate the model on the test set
    accuracy = knn.score(X_test, y_test)
    print(f'Test Set Accuracy: {accuracy}')

# Example usage
tune_knn_parameters(X, y)
```

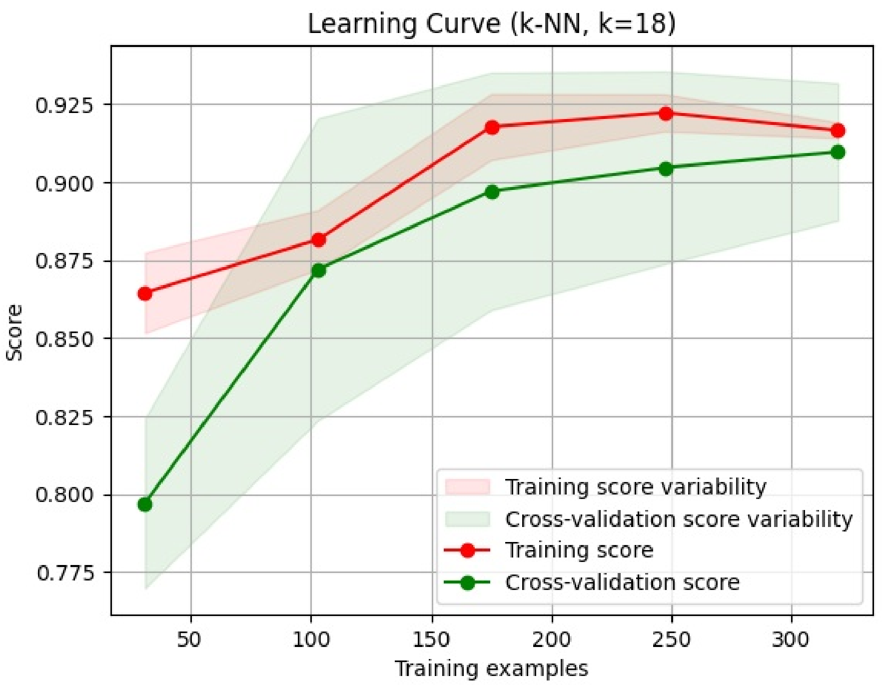
```
k=1, Mean Accuracy: 0.8998101265822784
k=2, Mean Accuracy: 0.8972784810126584
k=3, Mean Accuracy: 0.9148417721518987
k=4, Mean Accuracy: 0.9123734177215189
k=5, Mean Accuracy: 0.919873417721519
k=6, Mean Accuracy: 0.9223101265822784
k=7, Mean Accuracy: 0.9248417721518987
k=8, Mean Accuracy: 0.9298101265822785
k=9, Mean Accuracy: 0.9323417721518987
k=10, Mean Accuracy: 0.9298417721518988
k=11, Mean Accuracy: 0.9298417721518988
k=12, Mean Accuracy: 0.9298417721518988
k=13, Mean Accuracy: 0.9298417721518988
k=14, Mean Accuracy: 0.9223101265822784
k=15, Mean Accuracy: 0.9273101265822785
k=16, Mean Accuracy: 0.9248101265822786
k=17, Mean Accuracy: 0.9273101265822785
k=18, Mean Accuracy: 0.9248101265822786
k=19, Mean Accuracy: 0.9198101265822786
```

In [76]:
```
lazy_classification(X_train, y_train, X_test, y_test, 18)
```

```
Confusion Matrix:
[[58  2]
 [ 8 32]]

Classification Report:
              precision    recall  f1-score   support

           B       0.88      0.97      0.92        60
           M       0.94      0.80      0.86        40

    accuracy                           0.90       100
   macro avg       0.91      0.88      0.89       100
weighted avg       0.90      0.90      0.90       100

Test Set Accuracy: 0.9
```

In [77]:
```
import numpy as np

def plot_learning_curve(estimator, title, X, y, cv=5, n_jobs=-1, train_sizes=np.linspace(.1, 1.0, 5)):
    """
   Parameters:
- estimator: The entity employed to fit the data.
- title: The chart's title.
- X: Training vector, where n_samples represents the number of samples, and n_features signifies the number of
- y: Target corresponding to X for classification or regression.
- cv: Cross-validation generator or an iterable, default=None.
- n_jobs: The number of parallel jobs to execute, default=None.
- train_sizes: Relative or absolute quantities of training examples utilized for generating the learning curve.
```

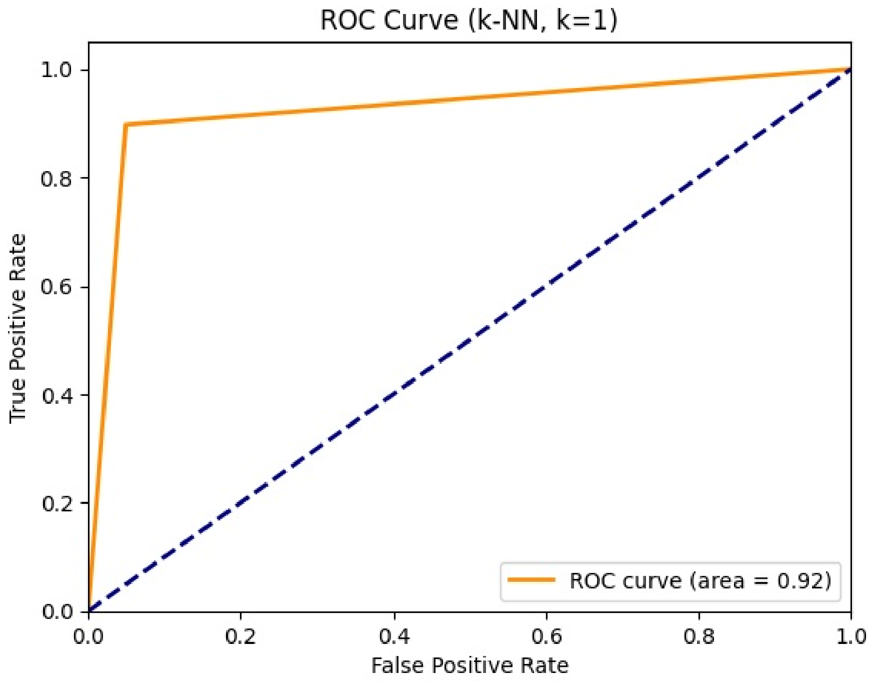
```
Returns:
- plt: Matplotlib plot object.

    """
    plt.figure()
    plt.title(title)
    plt.xlabel("Training examples")
    plt.ylabel("Score")

    train_sizes, train_scores, test_scores, fit_times, _ = \
        learning_curve(estimator, X, y, cv=cv, n_jobs=n_jobs, train_sizes=train_sizes, return_times=True)

    train_scores_mean = np.mean(train_scores, axis=1)
    train_scores_std = np.std(train_scores, axis=1)
    test_scores_mean = np.mean(test_scores, axis=1)
    test_scores_std = np.std(test_scores, axis=1)

    plt.grid()

    plt.fill_between(train_sizes, train_scores_mean - train_scores_std,
                     train_scores_mean + train_scores_std, alpha=0.1,
                     color="r", label="Training score variability")
    plt.fill_between(train_sizes, test_scores_mean - test_scores_std,
                     test_scores_mean + test_scores_std, alpha=0.1,
                     color="g", label="Cross-validation score variability")
    plt.plot(train_sizes, train_scores_mean, 'o-', color="r",
             label="Training score")
    plt.plot(train_sizes, test_scores_mean, 'o-', color="g",
             label="Cross-validation score")

    plt.legend(loc="best")
    return plt

# Assuming X_train, y_train, X_test, y_test are your training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Choose a k value for k-NN
k_value = 18
knn = KNeighborsClassifier(n_neighbors=k_value)

# Plot learning curve
plot_learning_curve(knn, f"Learning Curve (k-NN, k={k_value})", X_train, y_train, cv=5, n_jobs=-1)
plt.show()
```

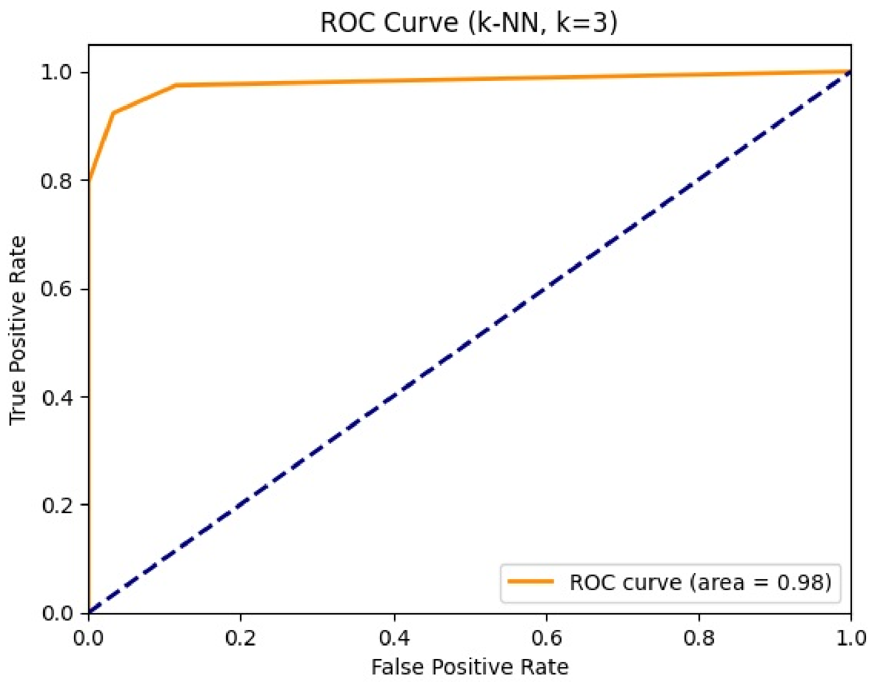


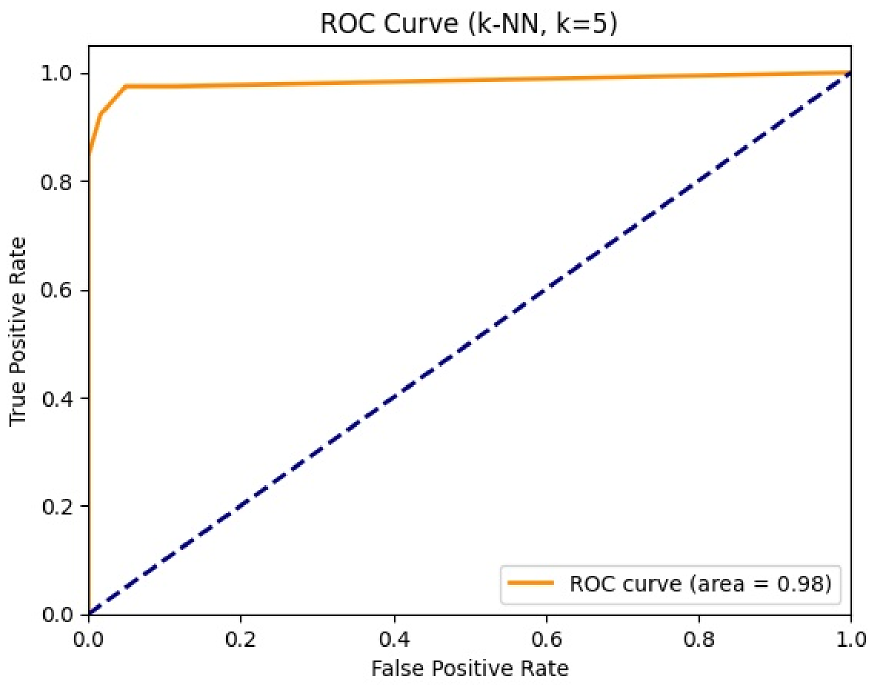
In [78]:

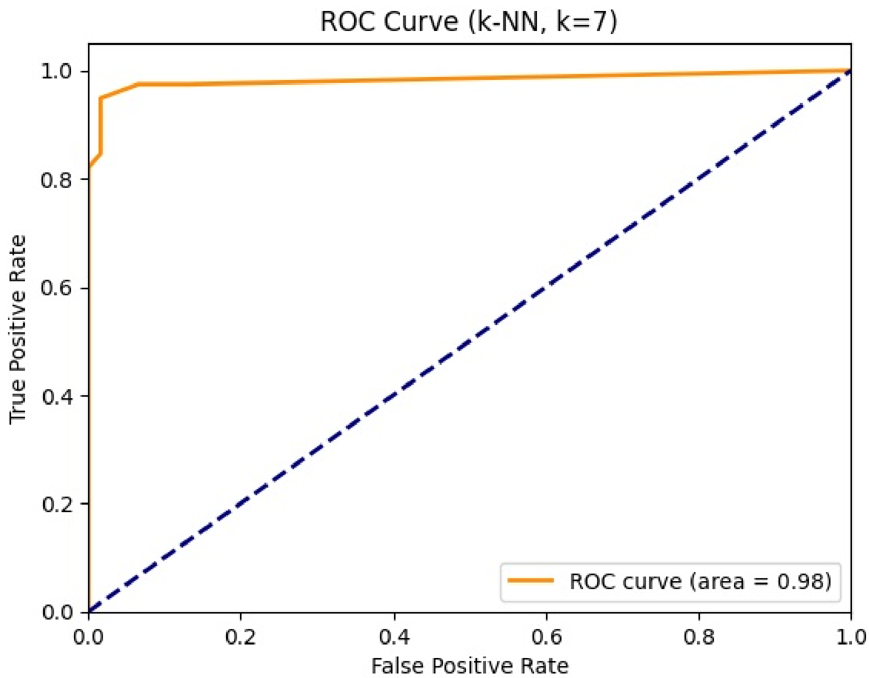
```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import roc_curve, auc
import matplotlib.pyplot as plt

# Assuming X, y are your dataset
# Replace this part with your actual data loading and preprocessing
# For example:
# df_heart_disease = pd.read_csv("your_dataset.csv")
# X = df_heart_disease.drop('target', axis=1)
# y = df_heart_disease['target']
```

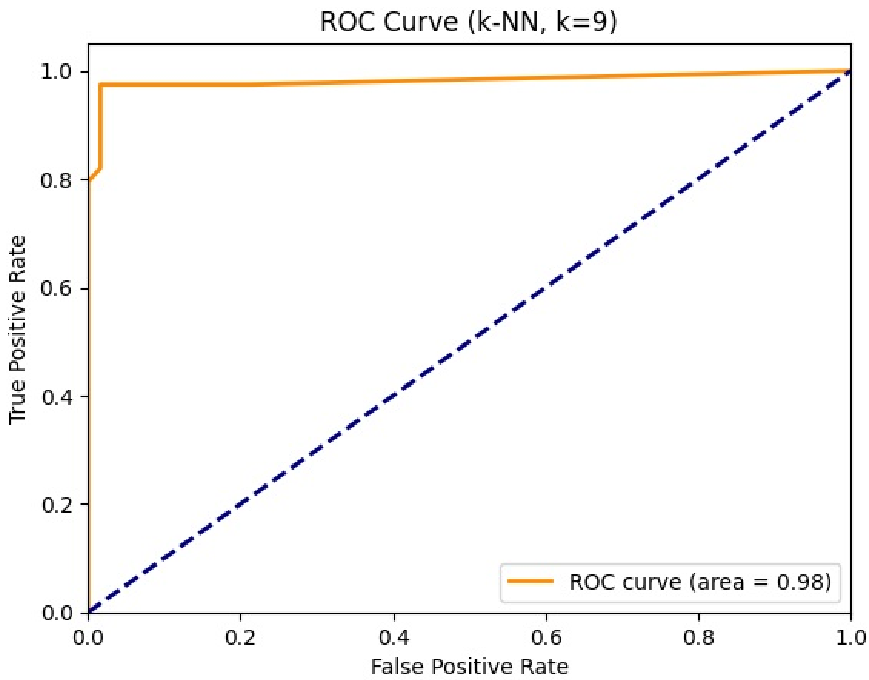
```python
# X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Function for tuning k-NN parameters
def tune_knn_parameters(X, y, k_value):
    # Split the data into training and testing sets
    X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

    # Train k-NN with the selected k
    knn = KNeighborsClassifier(n_neighbors=k_value)
    knn.fit(X_train, y_train)

    # Convert 'B' to 0 and 'M' to 1 in y_test
    y_test_binary = y_test.map({'B': 0, 'M': 1})

    # Plot ROC curve
    plot_roc_curve(knn, X_test, y_test_binary, f"ROC Curve (k-NN, k={k_value})")

# Function to plot ROC curve
def plot_roc_curve(model, X_test, y_test, title):
    y_prob = model.predict_proba(X_test)[:, 1]
    fpr, tpr, thresholds = roc_curve(y_test, y_prob)
    roc_auc = auc(fpr, tpr)

    plt.figure()
    plt.plot(fpr, tpr, color='darkorange', lw=2, label='ROC curve (area = {:.2f})'.format(roc_auc))
    plt.plot([0, 1], [0, 1], color='navy', lw=2, linestyle='--')
    plt.xlim([0.0, 1.0])
    plt.ylim([0.0, 1.05])
    plt.xlabel('False Positive Rate')
    plt.ylabel('True Positive Rate')
    plt.title(title)
    plt.legend(loc="lower right")
    plt.show()

# Assuming k_values is the list of k values you want to try
# For example, k_values = [1, 3, 5, 7, 9]
for k_value in [1, 3, 5, 7, 9]:
    tune_knn_parameters(X, y, k_value)
```




ROC Curve (k-NN, k=3)
True Positive Rate
False Positive Rate
ROC curve (area = 0.98)
0.0
0.2
0.4
0.6
0.8
1.0


ROC Curve (k-NN, k=5)
True Positive Rate
False Positive Rate
ROC curve (area = 0.98)
0.0
0.2
0.4
0.6
0.8
1.0





In [79]:

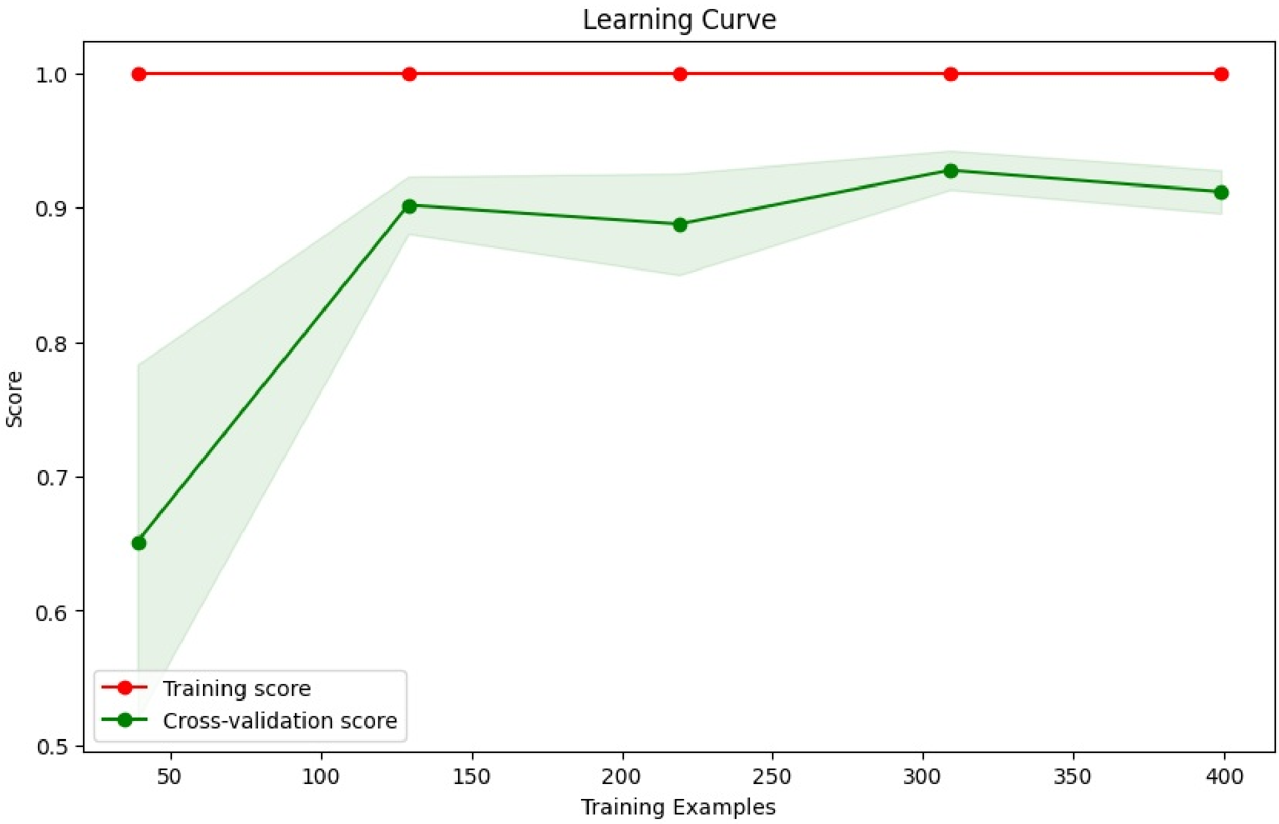
```
# Using Decision Tree Cfrom sklearn.tree import DecisionTreeClassifier

# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=40)

# Decision Tree Classifier
# Tuning the max depth parameter
best_depth = None
best_accuracy_tree = 0

# Add more values for tuning max depth
for depth_value in [None, 5, 10, 15]:
    # Train Decision Tree Classifier with the current max depth
    dtc = DecisionTreeClassifier(max_depth=depth_value, random_state=42)
    scores = cross_val_score(dtc, X_train, y_train, cv=5, scoring='accuracy')

    # Compute mean accuracy
```

```python
    mean_accuracy_tree = scores.mean()

    print(f"Max Depth={depth_value}, Mean Accuracy: {mean_accuracy_tree}")

    # Update best parameters if needed
    if mean_accuracy_tree > best_accuracy_tree:
        best_accuracy_tree = mean_accuracy_tree
        best_depth = depth_value

# Choose the optimal max depth based on the tuning results
optimal_depth = best_depth

# Perform classification with the optimal max depth
dtc = DecisionTreeClassifier(max_depth=optimal_depth, random_state=42)
dtc.fit(X_train, y_train)

# Predictions on the test set
y_pred_tree = dtc.predict(X_test)

# Decision Tree Classifier - Confusion Matrix
cm_tree = confusion_matrix(y_test, y_pred_tree)
print("\nDecision Tree Classifier - Confusion Matrix:")
print(cm_tree)

# Decision Tree Classifier - Classification Report
cr_tree = classification_report(y_test, y_pred_tree)
print("\nDecision Tree Classifier - Classification Report:")
print(cr_tree)

# Decision Tree Classifier - Evaluate the model on the test set
accuracy_tree = dtc.score(X_test, y_test)
print(f'Decision Tree Classifier - Test Set Accuracy: {accuracy_tree}')
```

```
Max Depth=None, Mean Accuracy: 0.934873417721519
Max Depth=5, Mean Accuracy: 0.9298417721518988
Max Depth=10, Mean Accuracy: 0.934873417721519
Max Depth=15, Mean Accuracy: 0.934873417721519

Decision Tree Classifier - Confusion Matrix:
[[57  3]
 [ 1 39]]

Decision Tree Classifier - Classification Report:
              precision    recall  f1-score   support

           B       0.98      0.95      0.97        60
           M       0.93      0.97      0.95        40

    accuracy                           0.96       100
   macro avg       0.96      0.96      0.96       100
weighted avg       0.96      0.96      0.96       100

Decision Tree Classifier - Test Set Accuracy: 0.96
```

In [80]:

```python
# Assuming `model` is your trained model and `X`, `y` are your features and labels
train_sizes, train_scores, test_scores = learning_curve(dtc, X, y, cv=5)

# Calculate mean and standard deviation across cross-validation folds
train_scores_mean = np.mean(train_scores, axis=1)
train_scores_std = np.std(train_scores, axis=1)
test_scores_mean = np.mean(test_scores, axis=1)
test_scores_std = np.std(test_scores, axis=1)

# Plot learning curve
plt.figure(figsize=(10, 6))
plt.fill_between(train_sizes, train_scores_mean - train_scores_std, train_scores_mean + train_scores_std, alpha
plt.fill_between(train_sizes, test_scores_mean - test_scores_std, test_scores_mean + test_scores_std, alpha=0.1
plt.plot(train_sizes, train_scores_mean, 'o-', color="r", label="Training score")
plt.plot(train_sizes, test_scores_mean, 'o-', color="g", label="Cross-validation score")

plt.title("Learning Curve")
plt.xlabel("Training Examples")
plt.ylabel("Score")
plt.legend(loc="best")
plt.show()
```



Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js