Solutions to Problem 1 of Hands-On Practice 1

Name: Anav Prasad (ap7152)   *Due: 11:59pm on Friday, July 15*

Collaborators:

Recorded results for testing Naive Bayes model on Hepatitis Dataset.

Solution:

Training Cycles = 200

- Replacing missing values with **minimum** values:

  accuracy: 76.00% +/– 8.99% (micro average: 76.13%)

	true 2	true 1	class precision
pred. 2	91	5	94.79%
pred. 1	32	27	45.76%
class recall	73.98%	84.38%	

- Replacing missing values with **maximum** values:

  accuracy: 70.17% +/– 9.32% (micro average: 70.32%)

	true 2	true 1	class precision
pred. 2	84	7	92.31%
pred. 1	39	25	39.06%
class recall	68.29%	78.12%	

- Replacing missing values with **average** values:

  accuracy: 74.71% +/– 7.60% (micro average: 74.84%)

	true 2	true 1	class precision
pred. 2	91	7	92.86%
pred. 1	32	25	43.86%
class recall	73.98%	78.12%	

- Replacing missing values with **zero** values:

  accuracy: 72.71% +/– 11.01% (micro average: 72.90%)

	true 2	true 1	class precision
pred. 2	86	5	94.51%
pred. 1	37	27	42.19%
class recall	69.92%	84.38%	

☐

Recorded results for testing Neural Net model on Hepatitis dataset:

Solution: Process Structure on RapidMiner:
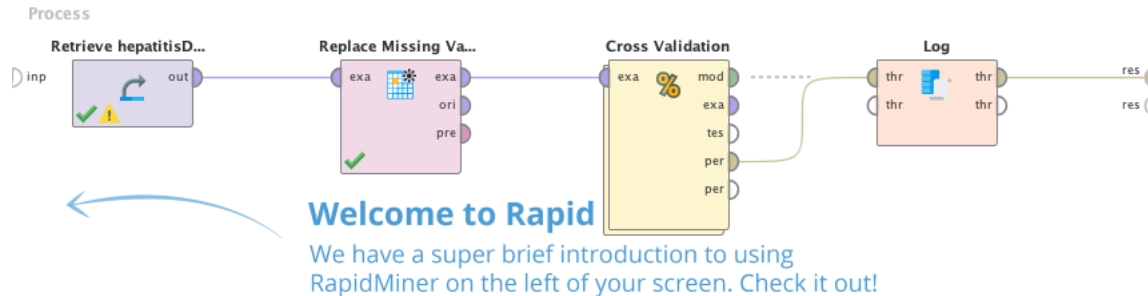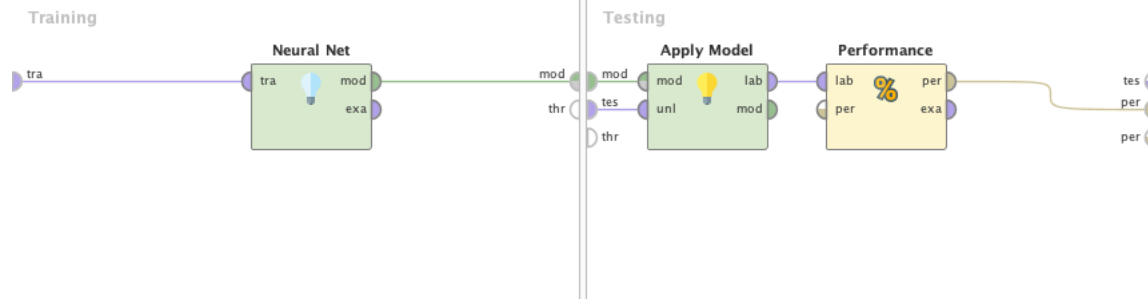
Figure 1: Overall Process

Figure 2: Cross Validation

Optimizer is optimizing two parameters: **Learning Rate**, and **Momentum**.
Results:

- Replacing missing values with *minimum* values; Training Cycles = 200:

accuracy: 83.92% +/- 5.31% (micro average: 83.87%)

	true 2	true 1	class precision
pred. 2	111	13	89.52%
pred. 1	12	19	61.29%
class recall	90.24%	59.38%	

- Replacing missing values with *maximum* values; Training Cycles = 200:

accuracy: 86.50% +/- 9.98% (micro average: 86.45%)

	true 2	true 1	class precision
pred. 2	114	12	90.48%
pred. 1	9	20	68.97%
class recall	92.68%	62.50%	

- Replacing missing values with *average* values; Training Cycles = 200:

accuracy: 85.92% +/- 6.40% (micro average: 85.81%)

	true 2	true 1	class precision
pred. 2	115	14	89.15%
pred. 1	8	18	69.23%
class recall	93.50%	56.25%	

- Replacing missing values with **zero** values; Training Cycles = 200:

accuracy: 84.00% +/- 7.40% (micro average: 83.87%)

	true 2	true 1	class precision
pred. 2	114	16	87.69%
pred. 1	9	16	64.00%
class recall	92.68%	50.00%	

As can be seen, best results were obtained by replacing missing values by maximum values. Trying to improve accuracy by increasing training cycles now:

- Replacing missing values with **maximum** values; Training Cycles = 500:

accuracy: 87.08% +/- 9.52% (micro average: 87.10%)

	true 2	true 1	class precision
pred. 2	114	11	91.20%
pred. 1	9	21	70.00%
class recall	92.68%	65.62%	

- Replacing missing values with **maximum** values; Training Cycles = 750:

accuracy: 86.46% +/- 9.23% (micro average: 86.45%)

	true 2	true 1	class precision
pred. 2	114	12	90.48%
pred. 1	9	20	68.97%
class recall	92.68%	62.50%	

- Replacing missing values with **maximum** values; Training Cycles = 1000:

accuracy: 86.46% +/- 9.02% (micro average: 86.45%)

	true 2	true 1	class precision
pred. 2	116	14	89.23%
pred. 1	7	18	72.00%
class recall	94.31%	56.25%	

Best Accuracy Obtained: **87.08%** □

Research and discuss very briefly, how would you deploy your algorithm built here (cite your sources)

Solution: There are several ways to deploy machine learning models.

- For web services: Flask, Streamlit, Docker or Kubernetes.

- For embedded models on edge devices: Using TensorFlow, Keras etc.

References:

- Deploy Machine Learning Models to Production With Flask, Streamlit, Docker, and Kubernetes on Google Cloud Platform

- https://towardsdatascience.com/3-ways-to-deploy-machine-learning-models-in-production-cdba15b00e

- https://stackoverflow.blog/2020/10/12/how-to-put-machine-learning-models-into-production/

$\square$