

Progress Report

Daniel Biro

January 26, 2021

Contents

1	Experiments - workflow version: 1.2.0; experiments revision: YYY	2
1.1	Methods	2
1.2	Results	2

1 Experiments - workflow version: 1.2.0; experiments revision: YYY

1.1 Methods

Genotyping was added with new scripts and models. These are all analogous to the existing ones, but produce differently shaped data (10x81 compared to the previous 4x81) and with the goal of predicting a nucleotide.

run_experiment.py has been updated to include calling different functions depending on the mode of execution (given by a new config item). The choice is between the previous scripts and the new version prefixed by "genotyping_". The Snakemake rules and bash scripts have been updated as well to call different scripts and produce differently named datasets based on the mode.

For genotyping the plotting and aggregation happens across all nucleotides predicted, but for the class prediction they have now been separated by class and more plots are produced to observe how the metrics for individual classes change over time. The new plotting seems to produce some unplausible plots on the simple GRU's training data but is working fine for the other models (the metrics are recorded in the same way and seem fine).

The GitHub repository has been updated but the documentation haven't been polished yet.

1.2 Results

I have noticed that at some point my Nvidia drivers broke so the model trainings progressed significantly slower when I last recorded the runtimes. After fixing things I have remeasured everything and made new measurements for the genotyping.

Approximate runtimes per 1000 epochs trained for the different model types for classification are: basic model types - 25mins, bi-directional models - 50mins, Transformer network (1 encoder layers) - 2.5hrs, Perceptron - 5mins.

For genotyping: basic model types - 20mins, bi-directional models - 50mins, Transformer network (1 encoder layers) - 3.5hrs, Perceptron - 6mins.

The previously best performing models were re-trained with 2000 epochs of training time (2x more than before), and the analogous models for genotyping have also been trained for 2000 epochs. The following tables present their results on the validation (table 1-5 for classification, table 11 for genotyping) and test (table 6-10 for classification, table 12 for genotyping) sets. The Transformer models are now with 1 encoder layer, but note that for genotyping only 8 runs were performed as of the writing of this report. For the genotyping results AUCs are omitted as well to make the tables fit onto the page.

The averaged results of classification with 2000 epochs showed improvement on most metrics for most models over just 1000 epochs training. The Results Summary excel file contains the averaged tables for both as well as all results seen here.

Graphs for everything can also be found in the given experiment's result folder (model name + Long and Genotyping suffix for these experiments). The graphs shown here are the Bi-GRU + 5% Dropout and Perceptron ones for classification on the validation set (figure 1-2) and for genotyping on the training and validation set (figure 3-4)

MODEL	NORMAL	GERMLINE VARIANT	SOMATIC VARIANT
GRU	94.16% \pm 0.16	98.72% \pm 0.04	93.97% \pm 0.17
BI-GRU + 5% DROPOUT	94.58% \pm 0.06	98.88% \pm 0.04	94.44% \pm 0.05
TRANSFORMER + 10% DROPOUT	86.78% \pm 0.12	97.51% \pm 0.04	87.96% \pm 0.1
PERCEPTRON	95.32% \pm 0.03	99.17% \pm 0.0	94.9% \pm 0.03

Table 1: Class Accuracies of best models on the validation set

MODEL	NORMAL	GERMLINE VARIANT	SOMATIC VARIANT
GRU	97.89% \pm 0.24	99.06% \pm 0.11	89.1% \pm 0.18
BI-GRU + 5% DROPOUT	98.39% \pm 0.11	98.85% \pm 0.18	89.78% \pm 0.07
TRANSFORMER + 10% DROPOUT	80.55% \pm 0.23	99.57% \pm 0.15	88.04% \pm 0.1
PERCEPTRON	99.08% \pm 0.03	100% \pm 0.0	90.41% \pm 0.05

Table 2: Class Precisions of best models on the validation set

MODEL	NORMAL	GERMLINE VARIANT	SOMATIC VARIANT
GRU	87.03% \pm 0.22	91.38% \pm 0.25	99.35% \pm 0.16
BI-GRU + 5% DROPOUT	87.69% \pm 0.13	92.82% \pm 0.19	99.51% \pm 0.09
TRANSFORMER + 10% DROPOUT	87.56% \pm 0.11	81.95% \pm 0.33	86.11% \pm 0.19
PERCEPTRON	88.94% \pm 0.07	93.85% \pm 0.0	99.75% \pm 0.02

Table 3: Class Recalls of best models on the validation set

MODEL	NORMAL	GERMLINE VARIANT	SOMATIC VARIANT
GRU	92.14% \pm 0.21	95.06% \pm 0.17	93.94% \pm 0.16
BI-GRU + 5% DROPOUT	92.72% \pm 0.08	95.74% \pm 0.15	94.4% \pm 0.05
TRANSFORMER + 10% DROPOUT	83.91% \pm 0.12	89.9% \pm 0.19	87.07% \pm 0.11
PERCEPTRON	93.74% \pm 0.04	96.83% \pm 0.0	94.85% \pm 0.03

Table 4: Class F1 Scores of best models on the validation set

MODEL	NORMAL	GERMLINE VARIANT	SOMATIC VARIANT
GRU	0.9416 \pm 0.0044	0.9587 \pm 0.0024	0.9452 \pm 0.0036
BI-GRU + 5% DROPOUT	0.9384 \pm 0.0029	0.9662 \pm 0.0017	0.9514 \pm 0.0012
TRANSFORMER + 10% DROPOUT	0.8224 \pm 0.0034	0.8649 \pm 0.0013	0.8278 \pm 0.0006
PERCEPTRON	0.9572 \pm 0.0004	0.97 \pm 0.0008	0.9566 \pm 0.0007

Table 5: Class AUCs of best models on the validation set

MODEL	NORMAL	GERMLINE VARIANT	SOMATIC VARIANT
GRU	93.24% \pm 0.07	97.36% \pm 0.05	91.99% \pm 0.04
BI-GRU + 5% DROPOUT	93.36% \pm 0.05	97.47% \pm 0.04	92.1% \pm 0.03
TRANSFORMER + 10% DROPOUT	82.84% \pm 7.12	90.07% \pm 4.42	77.69% \pm 7.02
PERCEPTRON	93.43% \pm 0.01	97.22% \pm 0.0	91.7% \pm 0.01

Table 6: Class Accuracies of best models on the test set

MODEL	NORMAL	GERMLINE VARIANT	SOMATIC VARIANT
GRU	96.02% \pm 0.17	98.08% \pm 0.12	87.18% \pm 0.04
BI-GRU + 5% DROPOUT	96.11% \pm 0.15	98.48% \pm 0.12	87.32% \pm 0.04
TRANSFORMER + 10% DROPOUT	84.99% \pm 8.02	85.62% \pm 7.8	61.12% \pm 12.71
PERCEPTRON	95.84% \pm 0.04	99.67% \pm 0.02	86.78% \pm 0.02

Table 7: Class Precisions of best models on the test set

MODEL	NORMAL	GERMLINE VARIANT	SOMATIC VARIANT
GRU	84.12% \pm 0.07	84.13% \pm 0.33	98.46% \pm 0.05
BI-GRU + 5% DROPOUT	84.4% \pm 0.07	84.5% \pm 0.3	98.49% \pm 0.03
TRANSFORMER + 10% DROPOUT	84.81% \pm 0.78	73.24% \pm 6.52	69.28% \pm 14.34
PERCEPTRON	84.86% \pm 0.04	81.83% \pm 0.04	98.4% \pm 0.02

Table 8: Class Recalls of best models on the test set

MODEL	NORMAL	GERMLINE VARIANT	SOMATIC VARIANT
GRU	89.68% \pm 0.1	90.57% \pm 0.21	92.48% \pm 0.04
BI-GRU + 5% DROPOUT	89.88% \pm 0.07	90.95% \pm 0.16	92.57% \pm 0.03
TRANSFORMER + 10% DROPOUT	82.44% \pm 5.35	74.0% \pm 6.61	64.9% \pm 13.45
PERCEPTRON	84.86% \pm 0.04	81.83% \pm 0.04	92.22% \pm 0.01

Table 9: Class F1 Scores of best models on the test set

MODEL	NORMAL	GERMLINE VARIANT	SOMATIC VARIANT
GRU	0.9301 \pm 0.003	0.9298 \pm 0.0039	0.9206 \pm 0.0028
BI-GRU + 5% DROPOUT	0.9244 \pm 0.0033	0.9326 \pm 0.0045	0.9208 \pm 0.0016
TRANSFORMER + 10% DROPOUT	0.8769 \pm 0.0433	0.8146 \pm 0.0429	0.8231 \pm 0.0679
PERCEPTRON	0.9391 \pm 0.0005	0.9218 \pm 0.0013	0.9235 \pm 0.0004

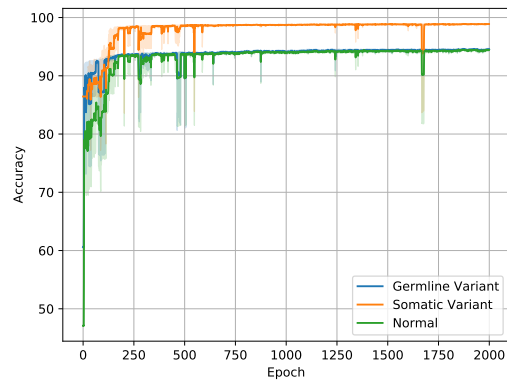
Table 10: Class AUCs of best models on the test set

MODEL	ACCURACY	PRECISION	RECALL	F1 SCORE
GRU	65.1% \pm 0.19	29.94% \pm 0.4	29.6% \pm 0.34	29.77% \pm 0.37
BI-GRU + 5% DROPOUT	65.59% \pm 0.29	31.01% \pm 0.58	30.78% \pm 0.56	30.78% \pm 0.56
TRANSFORMER + 10% DROPOUT	90.75% \pm 0.09	89.39% \pm 0.16	85.21% \pm 0.21	87.25% \pm 0.14
PERCEPTRON	99.42% \pm 0.03	98.86% \pm 0.05	98.82% \pm 0.06	98.84% \pm 0.06

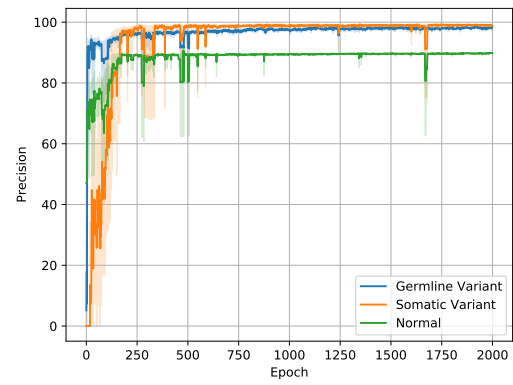
Table 11: Genotyping performance of best models on the validation set. Transformer model was only run 8 times

MODEL	ACCURACY	PRECISION	RECALL	F1 SCORE
GRU	64.27% \pm 0.09	28.38% \pm 0.18	28.29% \pm 0.18	28.33% \pm 0.18
BI-GRU + 5% DROPOUT	64.22% \pm 0.07	28.32% \pm 0.14	28.28% \pm 0.13	28.3% \pm 0.13
TRANSFORMER + 10% DROPOUT	83.53% \pm 6.19	77.24% \pm 9.51	75.8% \pm 7.21	76.5% \pm 8.47
PERCEPTRON	88.51% \pm 0.03	77.42% \pm 0.06	76.82% \pm 0.06	77.12% \pm 0.06

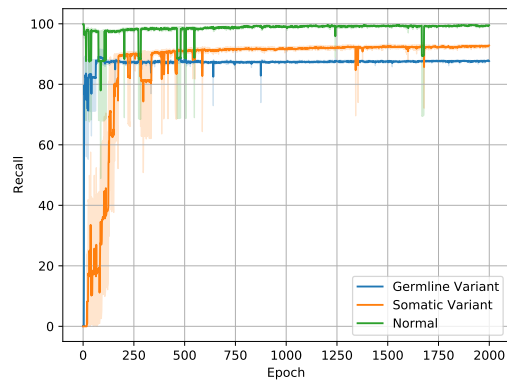
Table 12: Genotyping performance of best models on the test set. Transformer model was only run 8 times



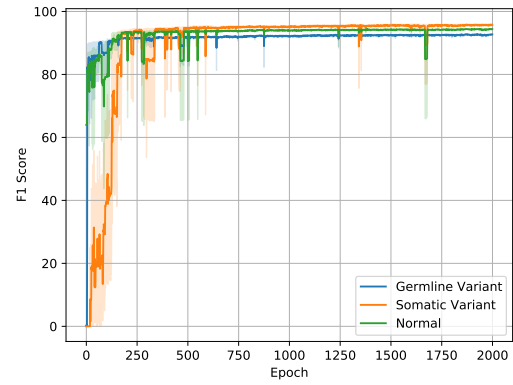
(a)



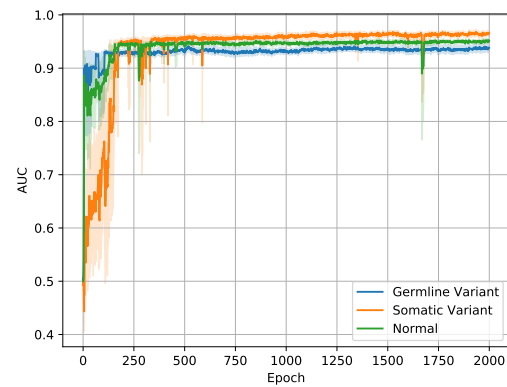
(b)



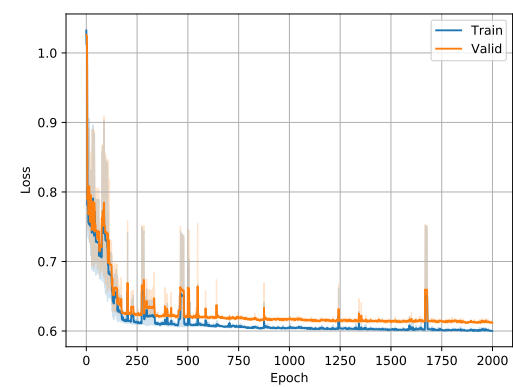
(c)



(d)

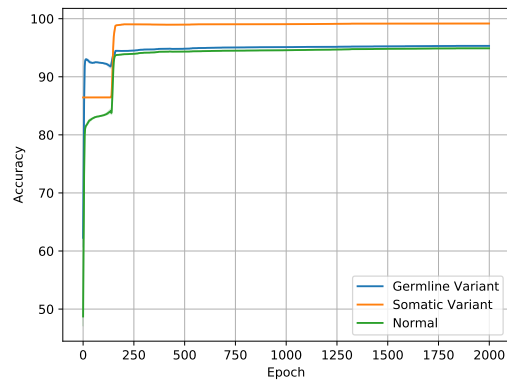


(e)

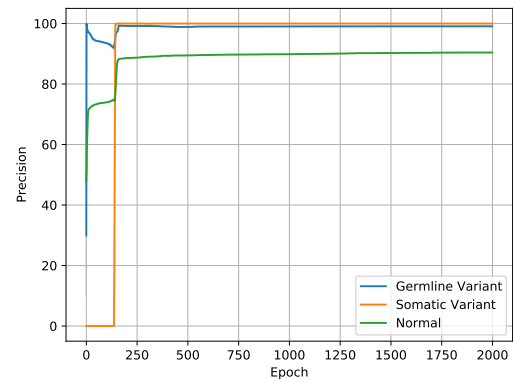


(f)

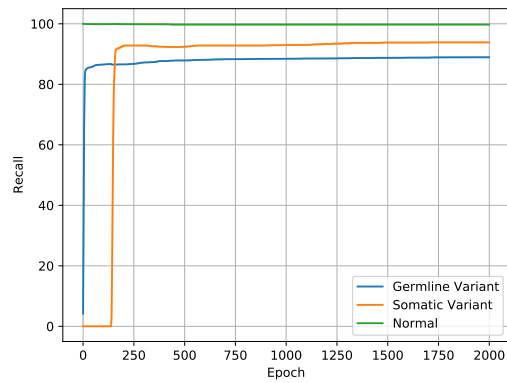
Figure 1: Figures for Bi-GRU + 5% Dropout model for classification on the validation set



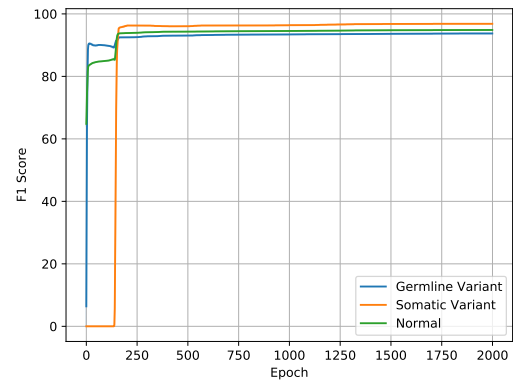
(a)



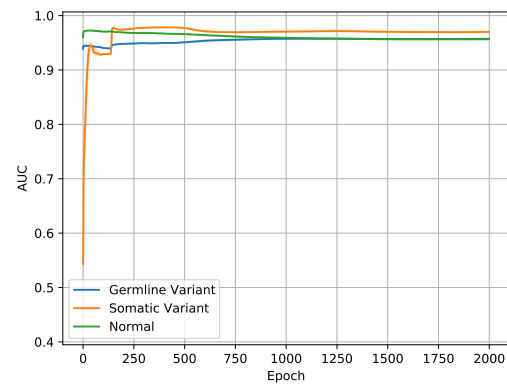
(b)



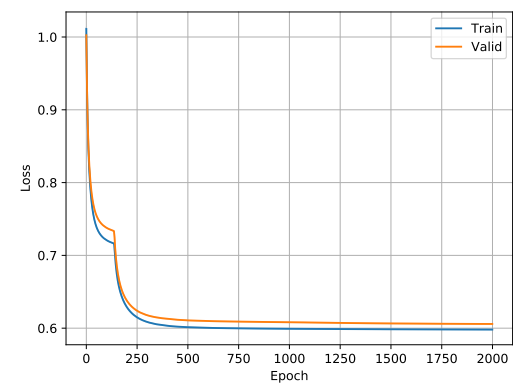
(c)



(d)

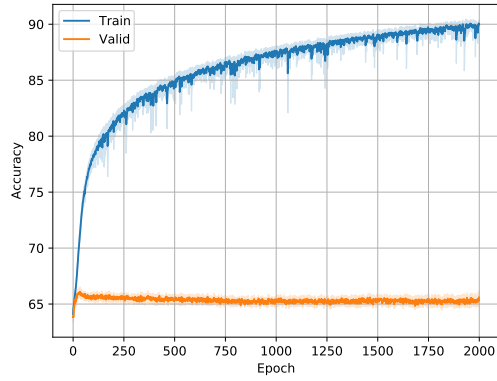


(e)

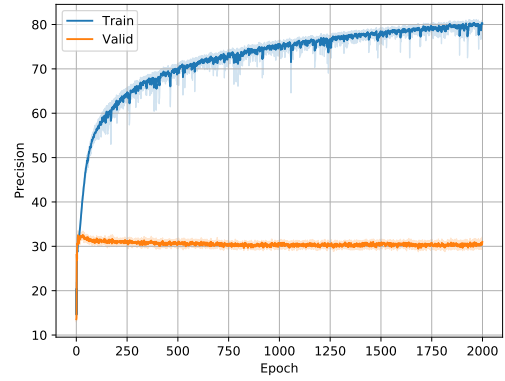


(f)

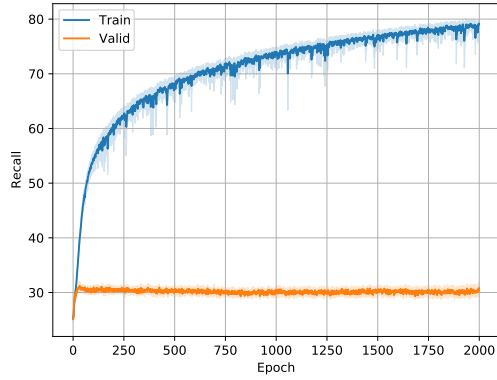
Figure 2: Figures for Perceptron model for classification on the validation set



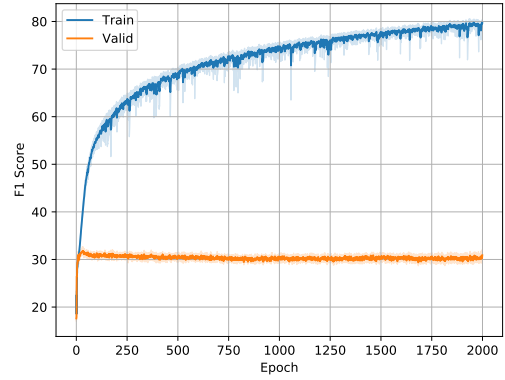
(a)



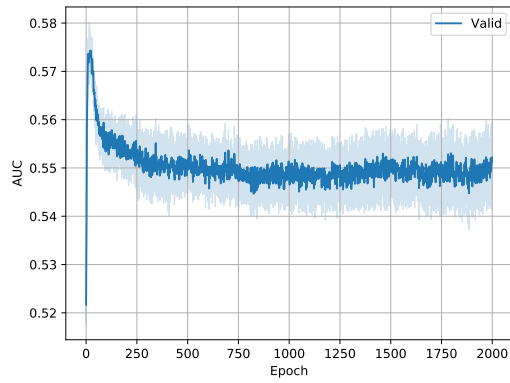
(b)



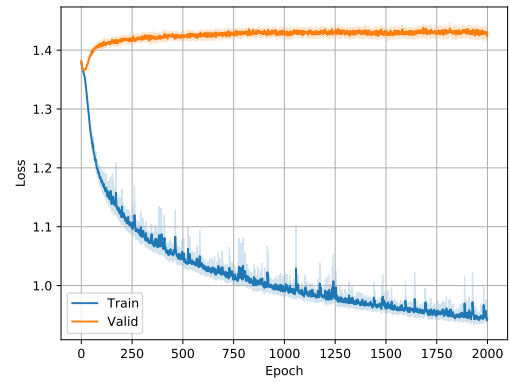
(c)



(d)

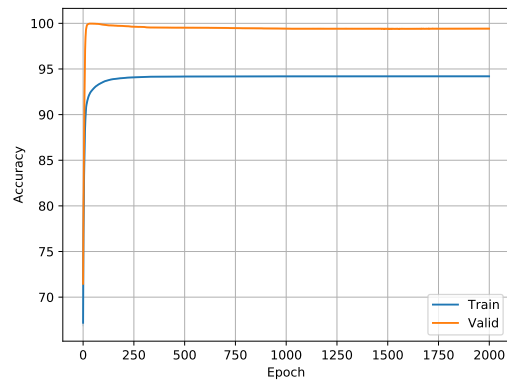


(e)

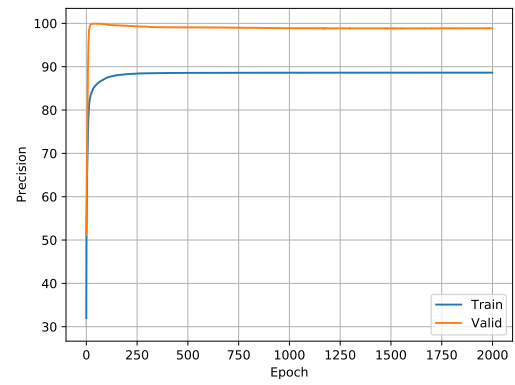


(f)

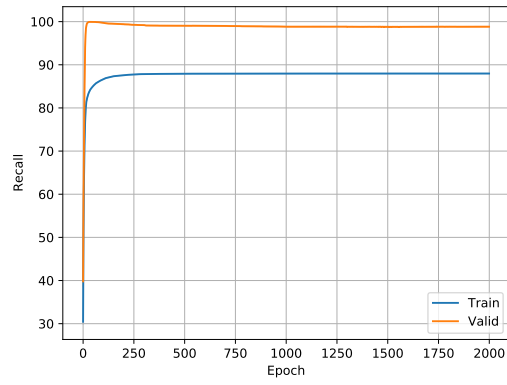
Figure 3: Figures for Bi-GRU + 5% Dropout model for genotyping



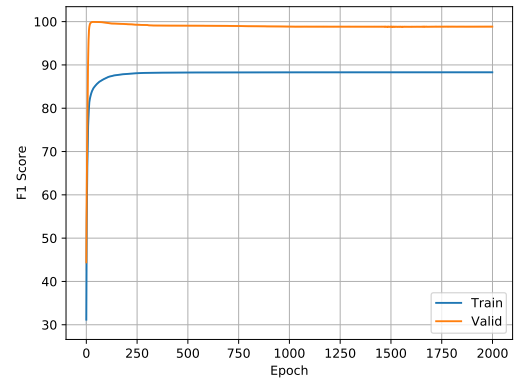
(a)



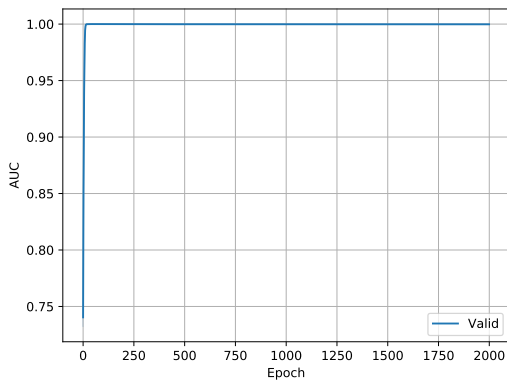
(b)



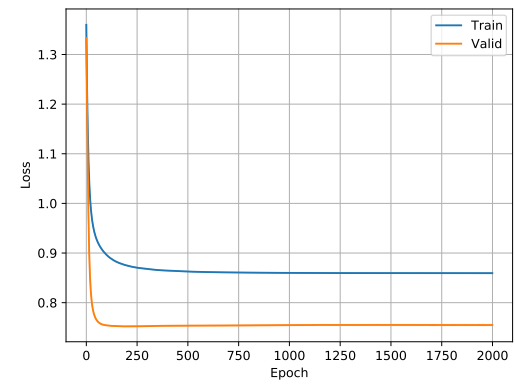
(c)



(d)



(e)



(f)

Figure 4: Figures for Perceptron model for genotyping