Skoltech
Skolkovo Institute of Science and Technology

Thesis Changes Log

Name of Candidate: Satyarth Mishra Sharma

PhD Program: Computational and Data Science and Engineering

Title of Thesis: Predicting Phenotype from Genotype with Machine Learning

Supervisor: Prof. Dmitry Yarotsky

The thesis document includes the following changes in answer to the external review process.

1.1.1:

- Clarified which Mendelian law of inheritance is referred to.

1.1.3:

- Rephrased duplicated sentence on Fisher's quantitative modelling of heridity to emphasize the difference from the Mendelian approach.

1.1.4:

- Replaced the erroneous claim that eye color is a monogenic trait.
- Added examples and supporting references for the polygenic nature of non-communicable diseases with a major health impact.

1.2.1:

- Clarified that SNPs are the most common form of genetic variation in terms of occurrence, not absolute quantity of varying genetic material.
- Clarified that the quoted frequency of SNPs relates to their occurrence when compared to a reference genome.
- Added a note of the impact of SNPs in non-coding regions on phenotypes by regulating gene expression.
- Replace unclear assertion that whole-genome sequencing captures an 'order of magnitude more data' with a mention of genome coverage as well as the types of genetic information captured.

2.2.8:

- Added tables with lists of covariates included in the analysis for each phenotype to the appendix.

3.1.3

- Updated notation of l_1 regularization to match rest of thesis.

3.2.1

- Added a note on the rationale for limiting the analysis to the FedAvg strategy rather than considering other federated learning strategies.

3.3.1

- Updated notation of r^2 score to match rest of thesis.

4.1.3

- Add a note on AlphaFold as a major breakthrough in the field and its limitations in predicting the consequences of mutations.

4.2.1

- Expanded the caption of Figure 4.1 to clarify the process of dimensionality reduction.

6.4

- Added a section with overall high level conclusions to tie the work together.