Здравствуйте!

Меня зовут Воронкина Дарья, я студентка группы ПМИ-81.

Я занимаюсь построением модели оценки времени дожития пациентов с БАС.

БАС - прогрессирующее, неизлечимое дегенеративное заболевание центральной нервной системы, приводящее к параличам и последующей атрофии мышц.

В 20% случаев болезнь связана с мутациями гена супероксиддисмутазы-1. Изучением мутаций в данном гене занимаются ученые из ИЦИГ, которые предоставили нам данные для построения модели. Данные содержат 182 признака водородных связей фермента SOD1 для 72 пациентов.

Так как признаки водородных связей представлены донорно-акцепторным взаимодействием фермента с описывающим его весом, то для каждого пациента можно рассматривать фермент в виде графа.

Строятся три различных варианта графа, отличающиеся в представлении вершин:

1) Вершина представлена в виде полной водородной связи;

2) Вершина представлена в виде позиции аминокислотного остатка с субъединицей водородной связи;

3) Вершина представлена в виде позиции аминокислотного остатка без субъединицы водородной связи.

Предполагается, что течение болезни влияет на структуру водородных связей в ферменте SOD1. Идея заключается в выявлении наличия зависимости между временем дожития пациента и изменением структуры фермента, поэтому рассматриваются следующие характеристики вершин графа.

Граф будет выглядеть след образом

Для каждого пациента строится граф.

Из вычисленных характеристик строятся обучающие выборки для регрессионной модели. Рассматриваемое уравнение выглядит след образом.

Число вычисленных характеристик вершин, выступающих объясняющими переменными, варьируется от 698 до 1068, что является проблемой, так как количество наблюдаемых значений равно 72. По этой же причине высока вероятность возникновения мультиколлинеарности между объясняющими переменными, что может привести к малоэффективному оцениванию параметров. Учитывая все особенности данных, задача решается методом частичных наименьших квадратов или partial least square (PLS). PLS регрессия выделяет небольшое количество латентных переменных, в пространстве которых связь между зависимой переменной и предикторами достигает максимального значения [2].

В процессе построения модели возникла проблема вырожденности обучающей выборки, на данный момент в качестве решение было решено выделить из графа фермента отдельные части, то есть подграфы и рассматривать их по отдельности. Таким образом предполагается решить проблему вырожденности и выделить наиболее значимые участки мутации фермента.

На данном этапе мною еще проводятся все необходимые вычисления храктеристик и подбор наилучшей модели.

Спасибо за внимание!