Здравствуйте с кем еще не здоровался. Тема у меня: регрессионные модели для определения повторного максимума в пауэрлифтинге.

Сначала я расскажу про сам пауэрлифтинг, затем – про актуальность исследования, опишу выборку и расскажу о результатах исследования.

Пауэрлифтинг – это силовой вид спорта, родственный тяжёлой атлетике, суть которого заключается в поднятии максимального веса в трёх упражнениях: приседание со штангой, жим лежа и становая тяги (всегда в таком порядке). Еще его называют силовым троеборьем, потому что 3 упражнения, а тяжелая атлетика – это силовое двоеборье (там рывок и толчок).

На соревновании выполняешь 3 упражнения, на каждое упражнение есть по 3 попытки, лучшие результаты по успешным попыткам суммируются, выигрывает атлет с наибольшей суммой в своей весовой категории – всё просто. Это настоящий спорт, ему уже лет 60, у него есть много спортивных федераций, можно сдавать нормативы мастера спорта и тому подобное.

Вес, который можешь поднять только на один раз, называется одноповторным максимумом. Если поднимаешь на несколько раз – это многоповторный максимум (двуповторный/трехповторный). Цель моего исследования – найти модель, которая предсказывает одноповторные максимумы.

Во-первых, зачем это нужно.

Вообще, очень полезно знать свой повторный максимум, не выясняя его напрямую, не делая проходку – так это называется, когда выясняешь, какой вес можешь на раз поднять. Это всегда очень травмоопасно, может привести к перетренированность, вдобавок это отдельная тренировка, к которой нужно циклически прийти, после нее запланировать спад – в общем, много рисков и траты времени.

А этот максимум нужно знать, чтобы

* Заказать оптимальный вес на соревнованиях. Если на соревнования поставишь меньше реального максимума, потеряешь преимущество. Если больше – потеряешь попытку, сильно устанешь, не сможешь показать лучшие результаты в следующих упражнениях. Или получишь травму, это тоже потеря преимущества на долгий срок.
* Чтобы правильно планировать программу тренировок. Очень многие программы тренировок, даже вообще не связанные с пауэрлифтингом, требуют тренироваться с весами в 80% от максимума, 70%, 50, а как узнать этот максимум человеку, который не занимался пауэрлифтингом или вообще спортом? Он еще не готов делать проходу, у него мышцы слабые, сухожилия слабые, техника не поставлена, такому человеку очень опасно поднимать большой для себя вес. Поэтому нужен способ узнать повторный максимум обходным путём.

Что вообще влияет на одноповторный максимум? То есть что влияет на то, какой вес человек сможет поднять на раз? Я не буду углубляться в физиологию: и так говорю очень много нематематики. Но основных компонент четыре. Это разные проявления генетики, опыт тренировок (чем человек больше тренируется, тем больше он развивает нужные качества), общая конституция – рост, вес, тип телосложения, и техника выполнения упражнения.

Какая гипотеза исследования?

Я ожидаю, что повторный максимум можно выразить примерно такой функциональной зависимостью. Это линейная компонента + нелинейная + ошибка. Здесь MRM это много повторный максимум, REPS – число повторений, x y – какие-то числа. Короче, осталось подобрать какую-то похожую зависимость. Как это сделать?

Это обычная задача обучения с учителем. Я много занимаюсь машинным обучением, в чем фишка обучения с учителем? Нужно предсказать какую-то переменную через другие переменные. Чтоб сделать это оптимально, нужно, чтобы в предикторах было побольше полезной информации и поменьше шумов. И нужно подобрать модель из правильного семейства, которая будет аппроксимировать реальную зависимость, подобрать у неё гиперпараметры и, в целом, всё. Конкретно здесь мы не можем учитывать многие вещи с предыдущего слайда, потому что уровень креатинфосфата и много другое нельзя узнать, не делая анализа крови и особых измерений. Но здесь есть одна переменная, которая собирает очень много полезной информации – это многоповторный максимум. Интуитивно понятно, что, зная, какой вес человек поднимет на три раза, можно неплохо оценить его результат на 1 раз. От этого и будем отталкиваться.

Про выборку.

Я провёл несколько опросов, использовал свои связи, привлек людей, в итоге за месяц собрал 174 наблюдения от более чем 150 человек. В основном это люди из России, но также там есть спортсмены из Болгарии, США и других стран. На картинках представлены некоторые графики по этой выборке…

Я очень подробно изучил выборку, провёл с ней много статистических тестов и обнаружил много значимых закономерностей, о которых догадывался. Например, на правом рисунке видно, что, в среднем, жим растёт с ростом индекса массы тела, но для приседа и тяги это верно только до которого порога. Это вполне очевидно: когда человек жмёт лежа, он лежит, ему его избыточный вес особо не мешает; а если он набрал его мышцами, то, наоборот, будет больше сила и больше жим. Например, сейчас в США есть спортсмен, Джулиус Мэддокс, он жмёт под 360 килограмм, но сам весит 200; до этого рекорд был у Сарычева (он, кстати, тоже участвовал в формировании выборки), он пожал почти 340 при весе в 170. Но когда человек приседает, там участвует еще и вес этого человека, поэтому большой вес становится проблемой.

На левом графике видно, что зависимость близка к линейной, но для каждого диапазона повторений будет как бы свой угол наклона. Я очень подробно расписал, почему так, но основная идея в том, что в разных диапазонах повторений участвуют свои режимы расхода энергии, поэтому разница есть.

В общем, могу сказать, что выборка достаточно репрезентативная, несмотря на размер.

Приближаемся к концу. Я прошёл долгий процесс подбора моделей. Исходил из тех, что были предложены до меня (не особо хорошие), а потом улучшал их, используя свою интуицию и свой опыт. Я, если что, 5 лет этим спортом занимался.

В итоге я получил такую линейную модель – и она очень хорошая. Здесь 4 коэффициента, 3 из них зависят от факторных переменных. Например, ActionB значит, что здесь стоит своё число в зависимости от движения: для жима своё, для приседа своё, для тяги своё.

Модель сама по себе значимая, статистически, каждое её слагаемое значимо, каждый уровень фактора тоже значимый, R2 очень близок к единице, ну и под все другие критерии подходит: остатки нормально распределены, гомоскедастичность, расстояния Кука небольшие и близкие и так далее. Ну и на кросс-валидации она явно превосходит остальные модели. И по качеству сама по себе модель очень неплохая.

Точно так же я подобрал нелинейную модель. Она основана на других паттернах, но зависит от тех же самых переменных и по качеству абсолютно такая же, как и линейная.

Здесь самое интересное – это то, от каких переменных идёт зависимость. От повторного максимума, числа повторений, типа упражнения и индекса массы тела, то есть роста и веса. Эти предикторы сами по себе настолько полезные, что не пришлось добавлять возраст, опыт тренировок, телосложение и другие доступные предикторы. И это – одно из основных интересных явлений в машинном обучении.

Даже если задача очень сложная, запутанная, непонятная, предсказать, сколько человек поднимет, или какая-то социология или ставки на спорт, выигрыш в игре, число голов – если взять несколько полезных признаков, похожая лёгкая формула часто будет работать. То есть большая часть на самом деле сложной закономерности часто объясняется подпространством низкой размерности. Живите теперь с этим.

Ну и заключение. Было проведено исследование, были найдены две модели хорошего качества, причём из разных семейств, и при этом для достижения не нужно было учитывать многие переменные. Также я запустил веб-приложение, на котором можно опробовать эти формулы на себе.