Школа анализа данных
Машинное обучение, часть 1
Теоретическое домашнее задание №2

Решите предложенные задачи. Решения необходимо оформить в виде PDF документа. Каждая задача должна быть подробно обоснована, задачи без обоснования не засчитываются. Решения пишутся в свободной форме, однако так, чтобы проверяющие смогли разобраться. Если проверяющие не смогут разобраться в решении какой-нибудь задачи, то она автоматически не засчитывается.

Задача 1 (0.5 балла) Кроссвалидация, LOO, k-fold.
Объясните, стоит ли использовать оценку leave-one-out-CV или k-fold-CV с небольшим k в случае, когда:

- обучающая выборка содержит очень малое количество объектов;
- обучающая выборка содержит очень большое количество объектов.

В случае с малым количеством объектов больше подходит LOO-CV, поскольку в этом случае увеличение размера обучающей выборки на 1 объект сильно влияет на качество модели. То есть при увеличении тестовой выборки на 1 качество модели падает сильнее, чем возрастает качество оценки этой модели. - уменьшается разброс между качеством на тесте и трейне, но и качество в среднем падает. Еще следует отметить, что поскольку объектов мало, то и обучить модель столько раз, сколько объектов в выборке можно за адекватное количество времени.

В случае с большим количеством объектов обучать модель столько раз, сколько объектов в выборке займет слишком много времени, поэтому LOO-CV не подходит. Если она очень большая, то даже KFold-CV будет работать слишком долго и лучше просто пользоваться HoldOut. Если нет, KFold-CV скорее всего успеет сойтись на $4/5$ объектов и среднее качество не будет сильно отличаться от качества при обучении на всех объектах, но при этом будет хорошая оценка на тестовом фолде, ведь в нем тоже много объектов.

Задача 2 (1.5 балла). Логистическая регрессия, решение оптимизационной задачи.

1. (0.5 балла) Докажите, что в случае линейно разделимой выборки не существует вектора параметров (весов), который бы максимизировал правдоподобие вероятностной модели логистической регрессии в задаче двухклассовой классификации.
   Пусть классы имеют метки $\{-1, 1\}$. Если выборка линейно разделима, это означает, что существует такое w', что $\forall i : y_i \langle w', x_i \rangle < 0$. Предположим, что существует w, максимизирующее правдоподобие вероятностной модели логистической регрессии, то есть для w достигается максимум

   $$L(w, X, y) = \sum_{i=1}^{N} \log \left(1 + e^{-y_i \langle w, x_i \rangle}\right)$$

   Но тогда если взять $w + w'$, правдоподобие увеличится - противоречие.

2. (0.3 балла) Предложите, как можно модифицировать модель, чтобы оптимум достигался.
Если в вероятностной модели предположить не только существование истинной зависимости между признаками и вероятностью положительного класса, но и априорное распределение на параметрах модели, объясняющееся неточностью измерений, представлений или наличием шума, то модель модифицируется таким образом:

$$p\left(X, Y, w; \sigma\right) = p\left(X, Y \mid w\right) p(w; \sigma)$$

Принцип максимума совместного правдоподобия данных и модели:

$$L_\sigma\left(w, X, Y\right) = \ln p\left(X, Y, w; \sigma\right) = \sum_{i=1}^{n} \ln p\left(x_i, y_i \mid w\right) + \ln p(w; \sigma) \to \max_w$$

Если предположить, что w имеет нормальное распределение $w \sim N\left(0, \sigma^2\right)$, где σ - гиперпараметр, то получаем:

$$\ln p(w; \sigma) = \ln\left(\frac{1}{(2\pi\sigma)^{n/2}} \exp\left(-\frac{\|w\|^2}{2\sigma}\right)\right) = -\frac{1}{2\sigma}\|w\|^2 + \text{const}(w)$$

Тогда L_σ - непрерывная по w функция, стремящаяся к минус бесконечности при $w \to \infty$, значит она достигает максимума.

3. (0.7 балла) Что можно сказать о единственности решения L2-регуляризованной задачи? Почему?
В случае L2- регуляризации логистической регрессии решение всегда единственно. Посмотрим на матрицу вторых производных:

$$\nabla L = -\frac{1}{\sigma}E - X^T D X$$

где D - матрица с положительными числами на диагонали. Тогда:

$$u^T \nabla L u = -\|u\|^2 - \|D^{1/2} X u\|^2 < 0$$

значит максимум единственный.

Задача 3 (0,5 балла). L^2-регуляризация.

Докажите, что L^2-регуляризованную линейную регрессию можно переписать в виде задачи наименьших квадратов для модифицированных данных.

Задача 4 (1.5 балла). L^1-регуляризация.

Рассмотрим задачу L^1-регуляризованной линейной регрессии, в которой ранг матрицы X меньше числа признаков D.

1. (0.5 балла) Докажите, что если у задачи более одного решения, то решений бесконечно много.
2. (0.5 балла) Докажите, что в этом случае для всех решений $\widehat{w}$ значение $X\widehat{w}$ одно и то же.
3. (0.5 балла) Докажите, что L^1-нормы всех решений $\widehat{w}$ одинаковы.

Задача 5 (1 балл). Обобщённая линейная модель.

Напомним, что гамма-распределение задаётся функцией плотности:

$$p(y \mid a, b) = \frac{1}{\Gamma(a) b^a} y^{a-1} e^{-\frac{y}{b}}$$

где $\Gamma(a)$ — гамма-функция

1. (0.3 балла) Докажите, что семейство гамма-распределений относится к экспоненциальному классу.

2. (0.7 балла) Как будет выглядеть соответствующая гамма-распределению обобщённая линейная модель? Найдите каноническую функцию связи и функционал, который надо оптимизировать, чтобы найти веса обобщённой линейной модели.

Задача 6 (4 балла) Обратное распространение ошибки.

В этой задаче вам нужно будет сделать простую вещь: написать формулы обратного распространения ошибки для нескольких слоёв. А чтобы вы лучше понимали, правильно ли у вас получилось, вам предстоит закодить своё решение и сдать в Яндекс.Контест (решение задачи не примем, если соответствующий слой не зайдёт в Контесте). Интерфейс слоёв точно такой же, как был на семинаре, и вы можете использовать все свои наработки с семинара. Вам предстоит одолеть:

1. **(0.5 балла)** LeakyReLU;

2. **(0.5 балла)** SoftPlus;

3. **(1 балл)** LogSoftMax;

4. **(1 балл)** нестабильную версию Negative log likelihood;

5. **(1 балл)** более стабильную версию Negative log likelihood.

На странице задания в ЛМС вас ждут:

- Ссылка на форму регистрации в соревнование;
- Ссылка на соревнование;
- Ноутбук `homework_part1_modules.ipynb`, который надо будет заполнить, преобразовать в .py-файл и сдать в систему;
- Ноутбук `homework_part1_test_modules.ipynb`, который поможет вам проверить локально, всё ли у вас ок.

Инструкция о том, как грузить решения в Контест, есть в соревновании. Не забывайте импортить numpy в сдаваемом .py-файле, без этого вас ждёт провал. Никакие специфические нейросетевые библиотеки использовать нельзя (да и не получится всё равно).

Обращаем внимание, что для успешной сдачи в Контест у вас должны также работать контейнер Sequential и линейный слой.

Задача 7 (0.5 балла) Нейронные сети.

Дана выборка из двух концентрических окружностей:

Допустим, что для классификации нужно обучить нейронную сеть — причем доступны только следующие слои: линейный $L(n, m)$ ($Wx + b,\ x \in \mathbb{R}^n,\ b \in \mathbb{R}^m$) и активация A (сигмоида или tanh), которые разрешено последовательно ставить друг после друга.

Вопрос: какие из приведенных ниже архитектур будут способны разделить выборку со 100% accuracy? Почему?

1. $L(2, 2) \to A \to L(2, 1)$

2. $L(2, 2) \to A \to L(2, 2) \to A \to L(2, 1)$

3. $L(2, 3) \to L(3, 1)$

4. $L(2,3) \to A \to L(3,1)$

5. $L(2,3) \to L(3,3) \to L(3,1)$

Задача 8 (0.5 балла) Нейронные сети, калибровка.

Глубокие нейронные сети часто являются плохо скалиброванными моделями. Что с ними не так? Почему?