Задача 1. Ridge Regression

Ridge regression – модель, идентичная обычной линейной регрессии, но с добавленной L_2 регуляризацией. Обозначим $X : n \times m$ матрицу признаковых описаний объектов, $Y : n$ – вектор значений целевой переменной, $\mathbf{w}$ – вектор весов (параметров модели), λ – параметр регуляризации. Покажите, что оптимальное $\mathbf{w}$ выражается как

$$\mathbf{w} = (XX^T + \lambda I)^{-1} XY$$

Задача 2. Softmax классификатор

Для того чтобы распространить логистическую регрессию на случай задачи классификации с m классами, смоделируем вероятность принадлежности к $k \in 1 \dots m$ классу с помощью softmax-функции

$$p(y_k|x) = \frac{\exp(a_k)}{\sum_{j=1}^{m} \exp(a_j)}, \quad a_k = \mathbf{w}_k^T \mathbf{x}$$

Для такой модели выпишите функцию правдоподобия и выражение для ее градиента.

Задача 3. SVM

Сопоставьте формулировки задач SVM (слева) и полученные разделяющие поверхности (справа). Аргументируйте свой ответ.

1.

$$\min \frac{1}{2}\mathbf{w} \cdot \mathbf{w} + C \sum_{i=1}^{n} \xi_i$$

s.t. $\forall i = 1, \cdots, n$:
$\xi_i \geq 0$
$(\mathbf{w} \cdot \mathbf{x_i} + b) y_i - (1 - \xi_i) \geq 0$
and $C = 0.1$.

2.

$$\min \frac{1}{2}\mathbf{w} \cdot \mathbf{w} + C \sum_{i=1}^{n} \xi_i \tag{1}$$

s.t. $\forall i = 1, \cdots, n$:
$\xi_i \geq 0$
$(\mathbf{w} \cdot \mathbf{x_i} + b) y_i - (1 - \xi_i) \geq 0$
and $C = 1$.

3.

$$\max \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \mathbf{K}(\mathbf{x_i}, \mathbf{x_j}) \tag{2}$$

s.t. $\sum_{i=1}^{n} \alpha_i y_i = 0$;
$\alpha_i \geq 0, \forall i = 1, \cdots, n$;
where $\mathbf{K}(\mathbf{u}, \mathbf{v}) = \mathbf{u} \cdot \mathbf{v} + (\mathbf{u} \cdot \mathbf{v})^2$.

4.

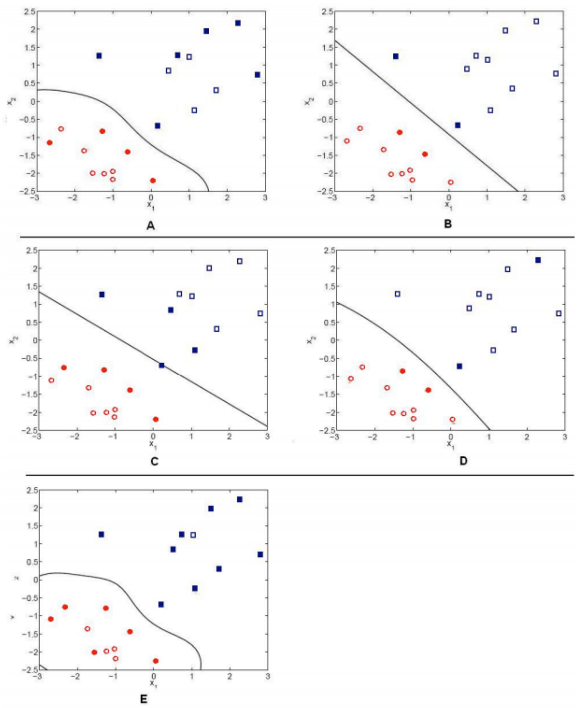
$$\max \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \mathbf{K}(\mathbf{x_i}, \mathbf{x_j}) \tag{3}$$

s.t. $\sum_{i=1}^{n} \alpha_i y_i = 0$;
$\alpha_i \geq 0, \forall i = 1, \cdots, n$;
where $\mathbf{K}(\mathbf{u}, \mathbf{v}) = \exp(-\frac{\|\mathbf{u}-\mathbf{v}\|^2}{2})$.

5.

$$\max \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \mathbf{K}(\mathbf{x_i}, \mathbf{x_j}) \tag{4}$$

s.t. $\sum_{i=1}^{n} \alpha_i y_i = 0$;
$\alpha_i \geq 0, \forall i = 1, \cdots, n$;
where $\mathbf{K}(\mathbf{u}, \mathbf{v}) = \exp(- \| \mathbf{u} - \mathbf{v} \|^2)$.



Задача 4. Naive Bayes

С помощью алгоритма Naive Bayes предскажите значение целевой переменной *buy_computer* для объекта со следующими значениями признаков:

```
age <= 30 & income = medium & student = yes & credit-rating = fair
```

Для обучения модели используйте данные из таблицы и сглаживане Лапласа.

RID	age	income	student	credit_rating	Class: buys_computer
1	<=30	high	no	fair	no
2	<=30	high	no	excellent	no
3	31 . . . 40	high	no	fair	yes
4	>40	medium	no	fair	yes
5	>40	low	yes	fair	yes
6	>40	low	yes	excellent	no
7	31 . . . 40	low	yes	excellent	yes
8	<=30	medium	no	fair	no
9	<=30	low	yes	fair	yes
10	>40	medium	yes	fair	yes
11	<=30	medium	yes	excellent	yes
12	31 . . . 40	medium	no	excellent	yes
13	31 . . . 40	high	yes	fair	yes
14	>40	medium	no	excellent	no

Задача 5. Решающие деревья

С помощью алгоритма CART постройте первые два уровня дерева решений на основании представленных данных. Используйте gini impurity.

```
medium   skiing   design     single  twenties no  -> highRisk
high     golf     trading    married forties  yes -> lowRisk
low      speedway transport  married thirties yes -> medRisk
medium   football banking    single  thirties yes -> lowRisk
high     flying   media      married fifties  yes -> highRisk
low      football security   single  twenties no  -> medRisk
medium   golf     media      single  thirties yes -> medRisk
medium   golf     transport  married forties  yes -> lowRisk
high     skiing   banking    single  thirties yes -> highRisk
low      golf     unemployed married forties  yes -> highRisk
```

Задача 6. Bias-Variance tradeoff

Рассмотрим N пар $(x_i, y_i) \in (\mathcal{X}, \mathcal{Y})$, выбранных независимо в соответствии с распределнием

$$x_i \sim h(x)$$

$$y_i = f(x_i) + \varepsilon_i$$

$$\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$$

Пусть функция, используемая для предсказания, линейна по y_i

$$\hat{f}(x_0) = \sum_{i=1}^{N} l_i(x_0, \mathcal{X}) y_i$$

Заметим, что веса $l_i(x_0, \mathcal{X})$ не зависят от y_i, но зависят от всей выборки $\mathcal{X}$.

1. Покажите, что KNN и линейная регрессия являются членами этого семейтсва моделей. Выпишите веса l_i в явном виде.
2. Выпишите среднеквадратичную ошибку

$$E_{\mathcal{Y}|\mathcal{X}}(f(x_0) - \hat{f}(x_0))^2$$

в виде суммы variance и квадрата bias.