Машинное обучение: линейные модели классификации и регрессии

Эмели Драль

CBR, Москва 2020

Программа курса

Курс состоит из **5ти** блоков:

1. **Базовые** концепции машинного обучения
2. **Линейные** модели классификации и регрессии
3. **Деревья решений** в классификации и регрессии, **ансамбли моделей**
4. Обучение **без учителя** и частичное обучение
5. **Нейронные сети** и глубокое обучение, backpropagation, регуляризация и методы оптимизации

План занятия

1. Обзор алгоритмов обучения с учителем
2. Линейные модели: интуиция
3. Линейные модели: построение
4. Обзор метрик качества (если успеем)

Обзор алгоритмов обучения с учителем

Обучение с учителем

Базовые концепты

Объекты и признаки:

- x – объект
- y – ответ
- $(f_1, f_2 \ldots f_n)$ – признаки,
- описывающие объекты

Модель:

- a: X -> Y
- a(x) = y
- A – семейство моделей

- X – пространство объектов
- Y – пространство ответов

Оценка качества

- Q(a, X) – ошибки модели a(x) на группе объектов X

Обучение с учителем

Как построить модель?

1. Подготовить набор данных $X = (x_i, y_i)_{i=1,l}$
2. Выбрать семейство алгоритмов A
3. Минимизировать ошибки модели Q(a , X) -> за счет этого получить конкретную модель a(x) из выбранного семейства A

Обучение с учителем

? Какие идеи заложены в модели?

Обучение с учителем

Семейства алгоритмов

- Метод ближайших соседей
- Линейные модели
- Деревья решений
- Ансамбли
- Нейронные сети

и др.

Обучение с учителем

Семейства алгоритмов

- Метод ближайших соседей: похожие объекты относятся к одному классу
- Линейные модели
- Деревья решений
- Ансамбли
- Нейронные сети

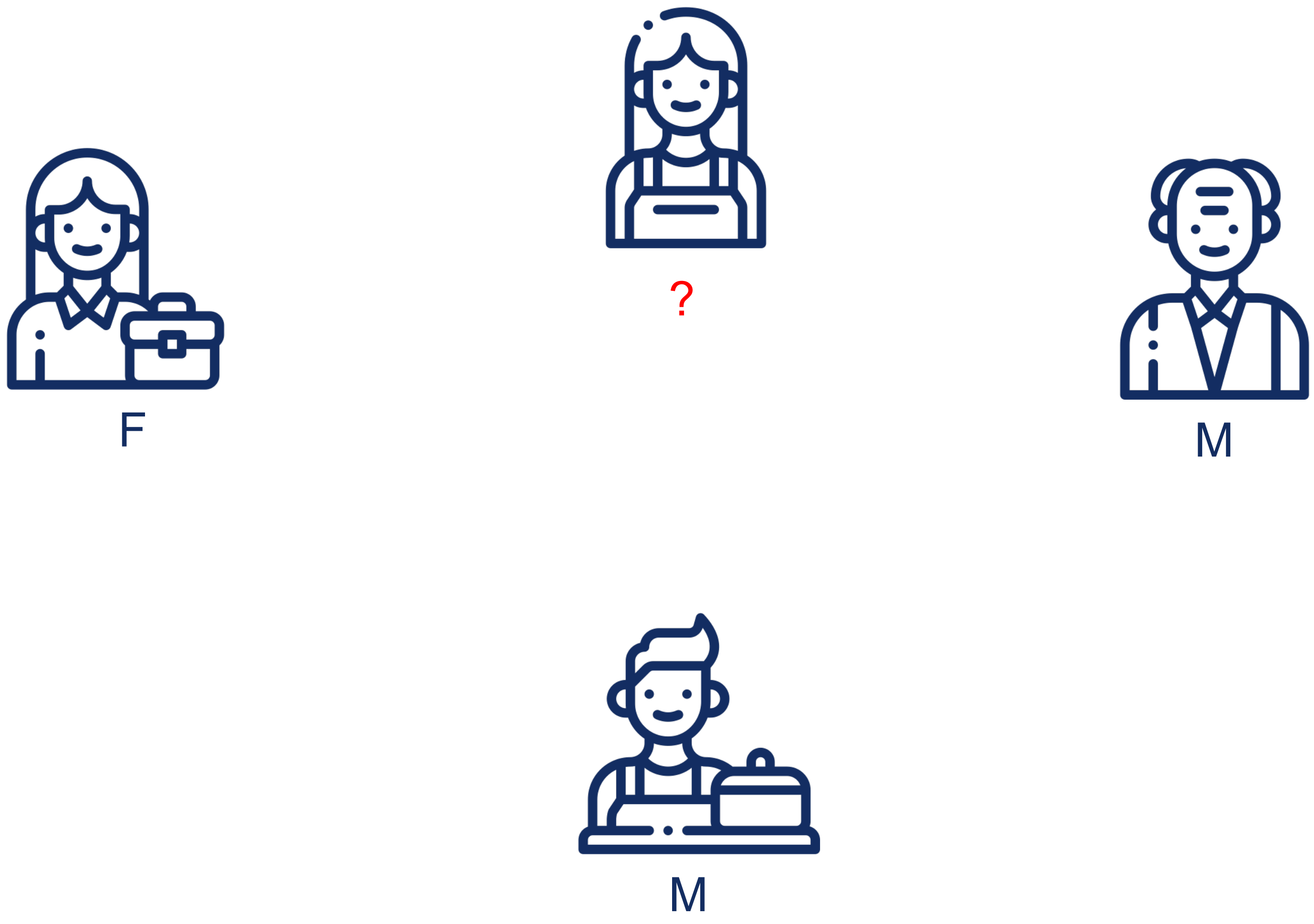
и др.

Обучение с учителем

Метод ближайших соседей

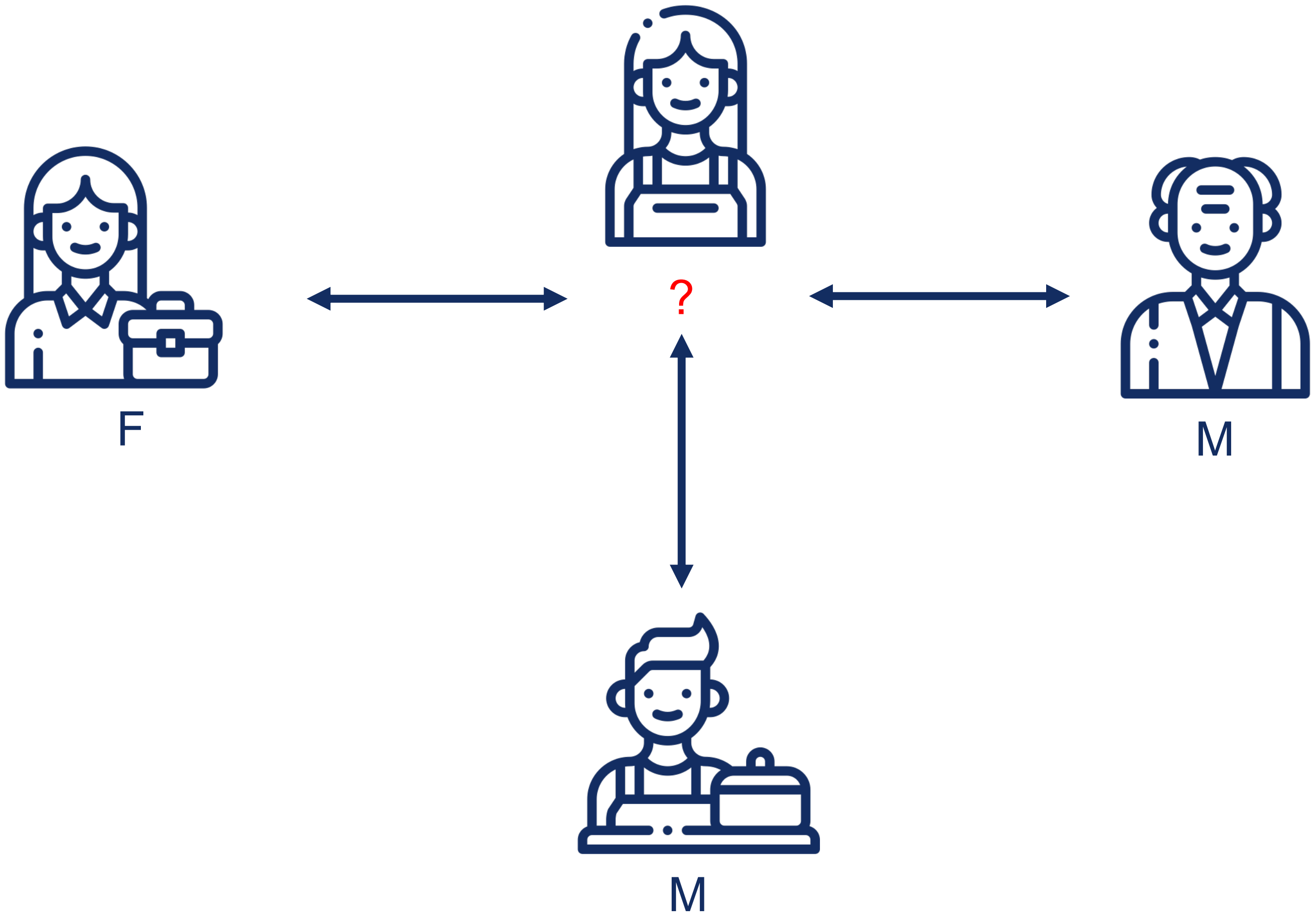
Обучение с учителем

Метод ближайших соседей



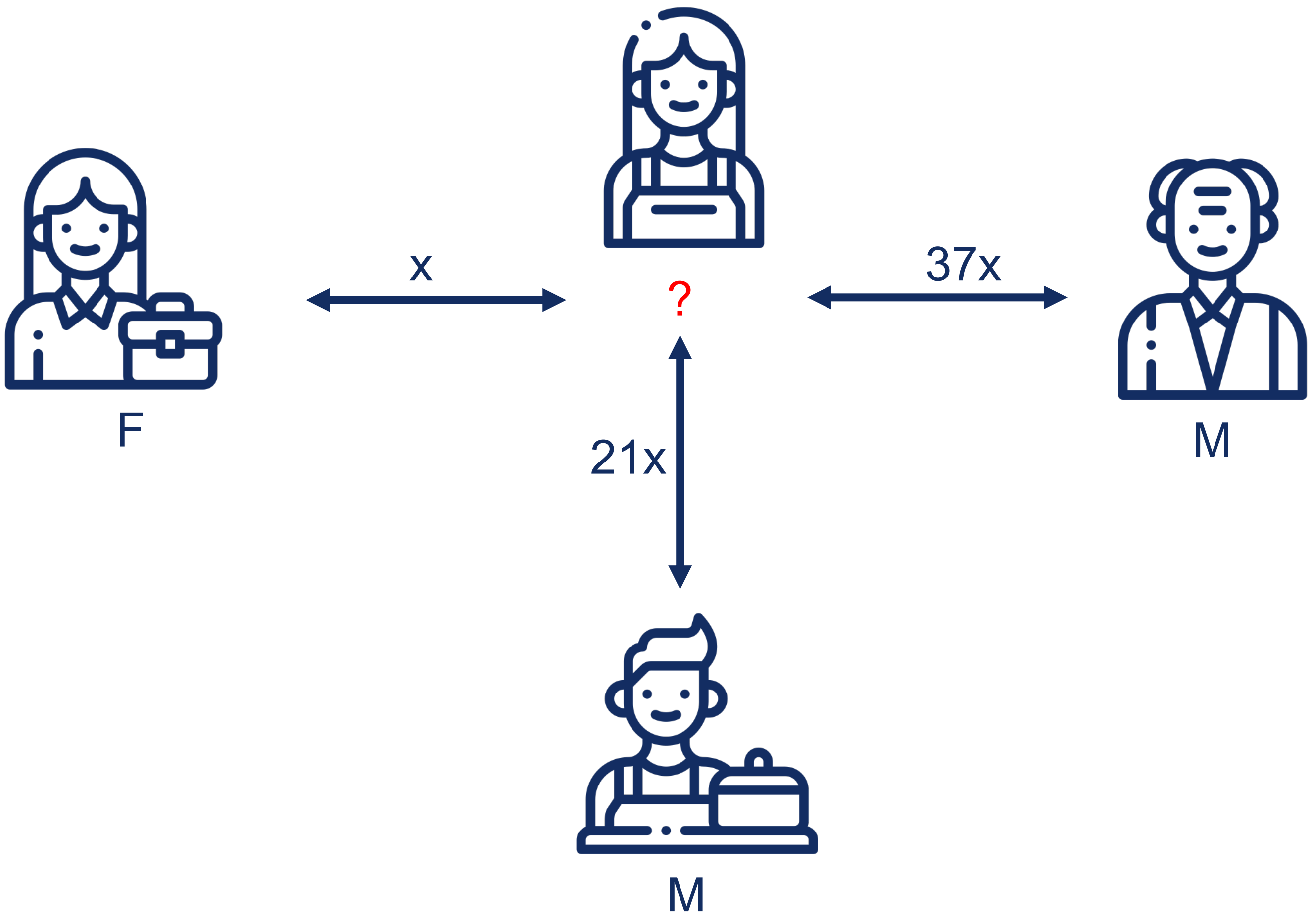
Обучение с учителем

Метод ближайших соседей



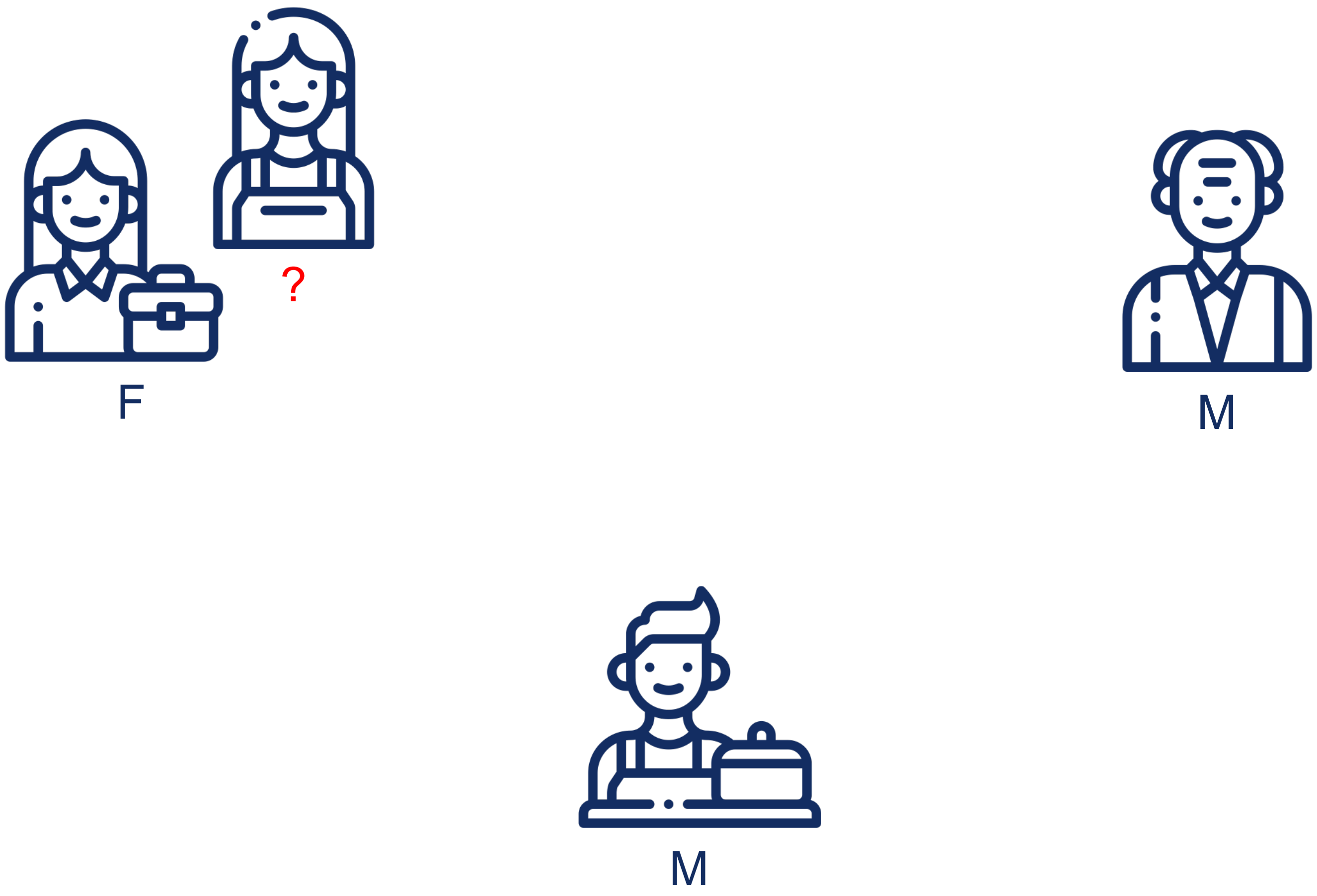
Обучение с учителем

Метод ближайших соседей



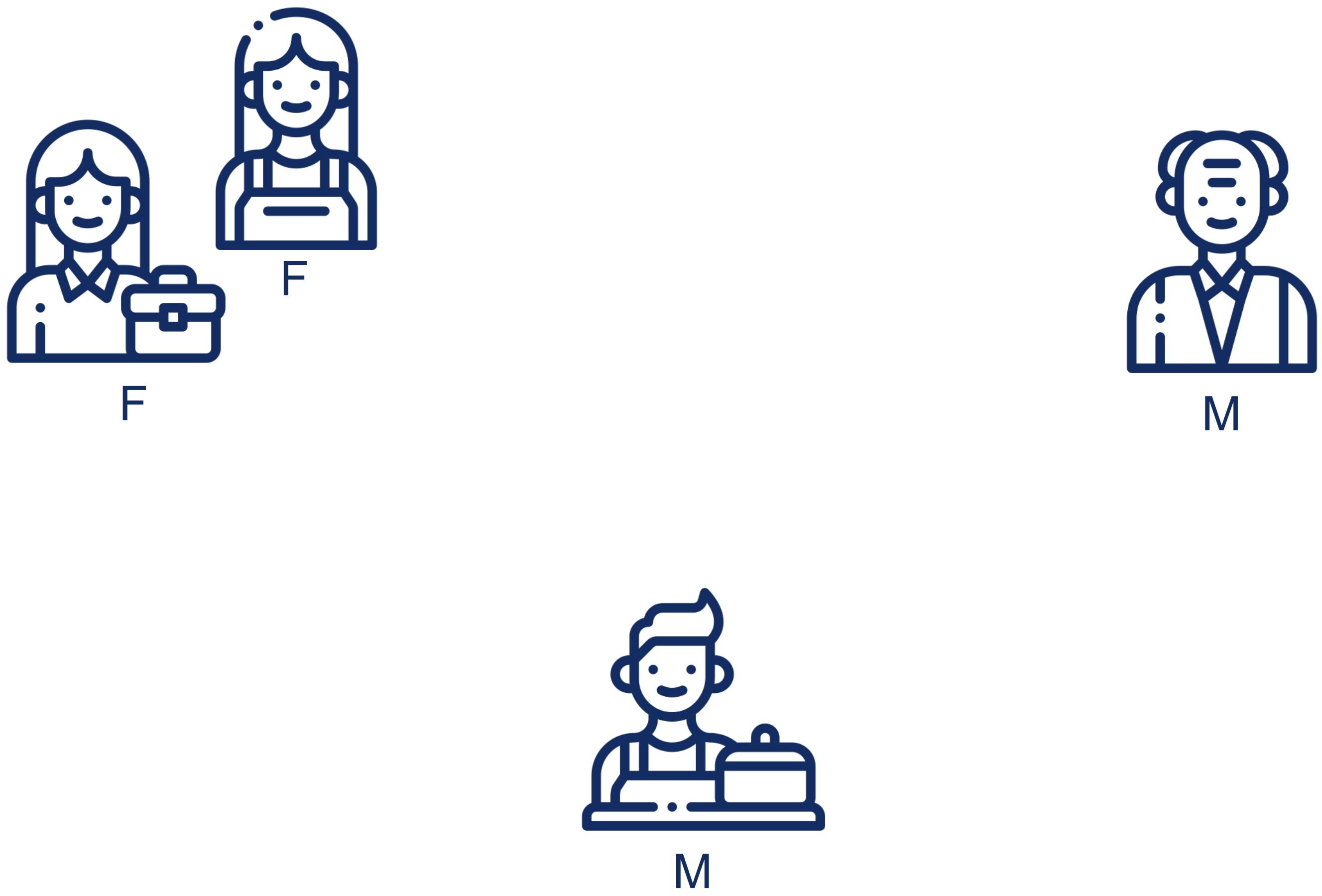
Обучение с учителем

Метод ближайших соседей



Обучение с учителем

Метод ближайших соседей



Обучение с учителем

Семейства алгоритмов

- Метод ближайших соседей: похожие объекты относятся к одному классу
- Линейные модели: класс линейно зависит от характеристик объекта
- Деревья решений
- Ансамбли
- Нейронные сети

и др.

Обучение с учителем

Линейная модель

Стоит ли одобрить кредит данному клиенту?

Обучение с учителем

Линейная модель

Стоит ли одобрить кредит данному клиенту?

Обучение с учителем

Линейная модель

Стоит ли одобрить кредит данному клиенту?

15

99

35

Обучение с учителем

Линейная модель

Стоит ли одобрить кредит данному клиенту?

 +  + 

15 99 35

Обучение с учителем

Семейства алгоритмов

- Метод ближайших соседей: похожие объекты относятся к одному классу
- Линейные модели: класс линейно зависит от характеристик объекта
- Деревья решений: класс получается в результате последовательных ответов на простые вопросы
- Ансамбли
- Нейронные сети

и др.

Обучение с учителем

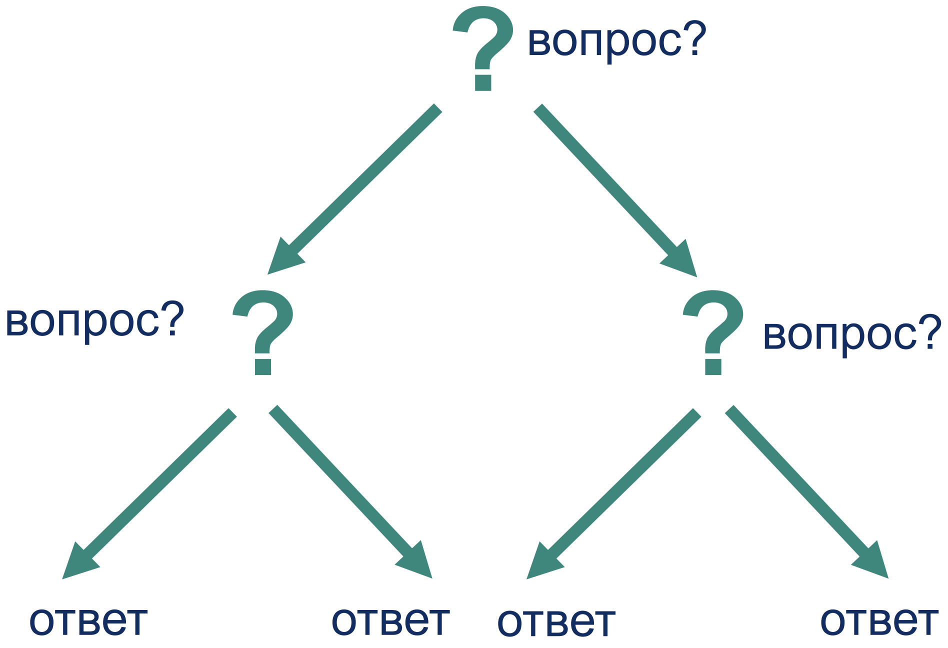
Дерево решений

Уйдет ли этот клиент к конкуренту?

Обучение с учителем

Дерево решений

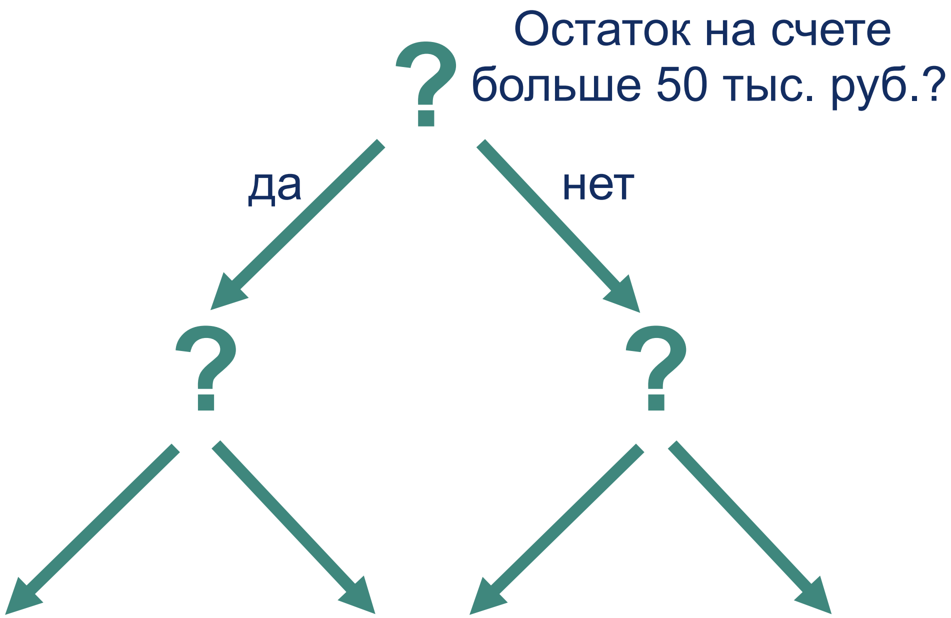
Уйдет ли этот клиент к конкуренту?



Обучение с учителем

Дерево решений

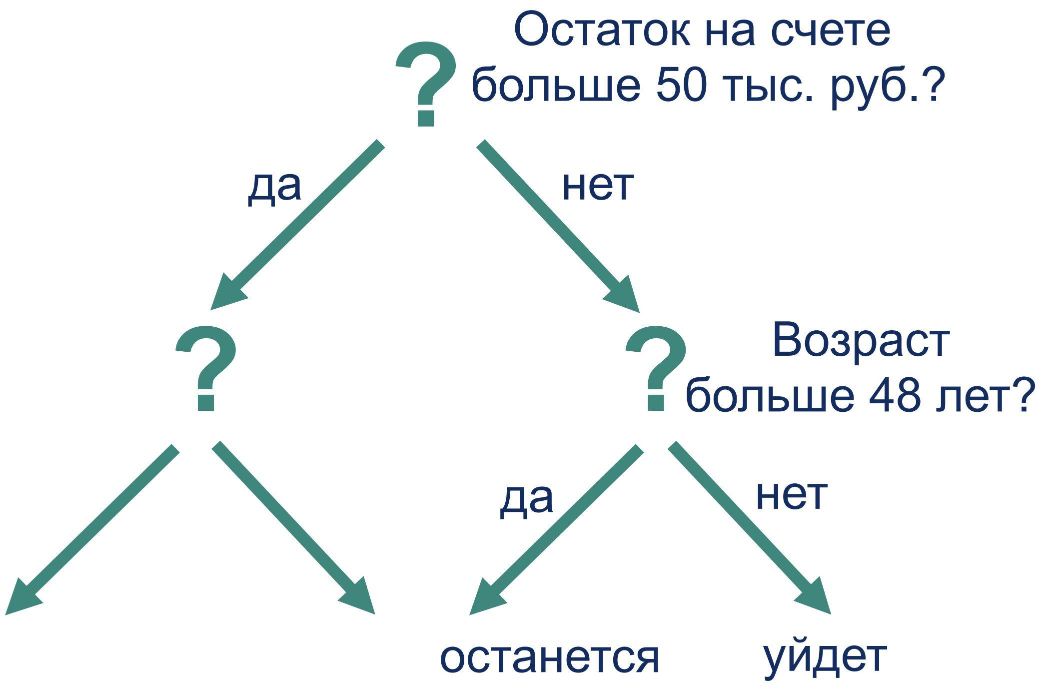
Уйдет ли этот клиент к конкуренту?



Обучение с учителем

Дерево решений

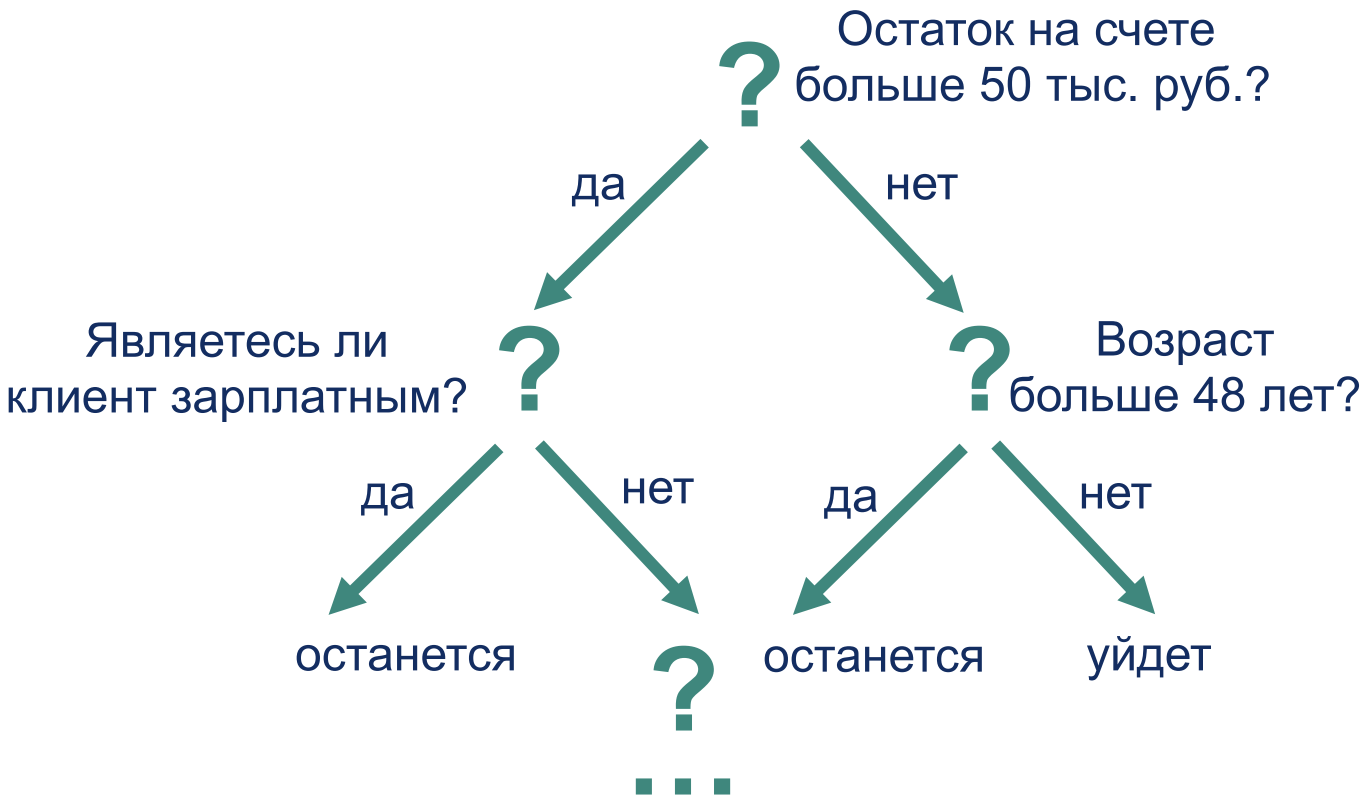
Уйдет ли этот клиент к конкуренту?



Обучение с учителем

Дерево решений

Уйдет ли этот клиент к конкуренту?



Обучение с учителем

Семейства алгоритмов

- Метод ближайших соседей: похожие объекты относятся к одному классу
- Линейные модели: класс линейно зависит от характеристик объекта
- Деревья решений: класс получается в результате последовательных ответов на простые вопросы
- Ансамбли: решение принимается на основе ответов нескольких моделей
- Нейронные сети

и др.

Обучение с учителем

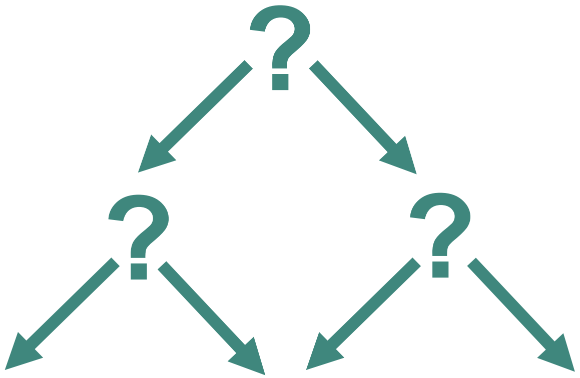
Ансамбли

Уйдет ли этот клиент к конкуренту?

Обучение с учителем

Ансамбли

Уйдет ли этот клиент к конкуренту?

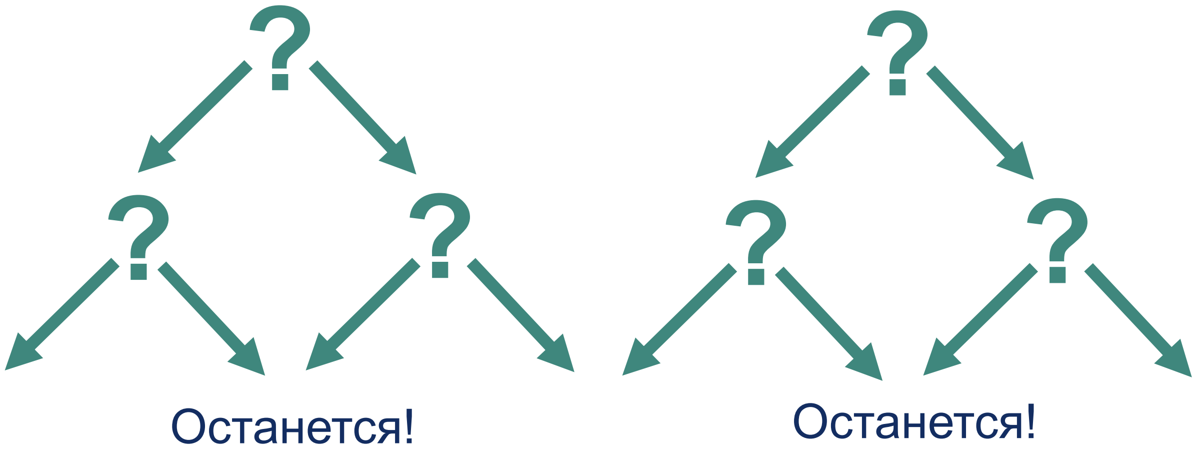


Останется!

Обучение с учителем

Ансамбли

Уйдет ли этот клиент к конкуренту?



Обучение с учителем

Ансамбли

Уйдет ли этот клиент к конкуренту?

Обучение с учителем

Ансамбли

Уйдет ли этот клиент к конкуренту?

2 vs 1

Обучение с учителем

Семейства алгоритмов

- Метод ближайших соседей: похожие объекты относятся к одному классу
- Линейные модели: класс линейно зависит от характеристик объекта
- Деревья решений: класс получается в результате последовательных ответов на простые вопросы
- Ансамбли: решение принимается на основе ответов нескольких моделей
- Нейронные сети: нелинейная комбинация линейных моделей

и др.

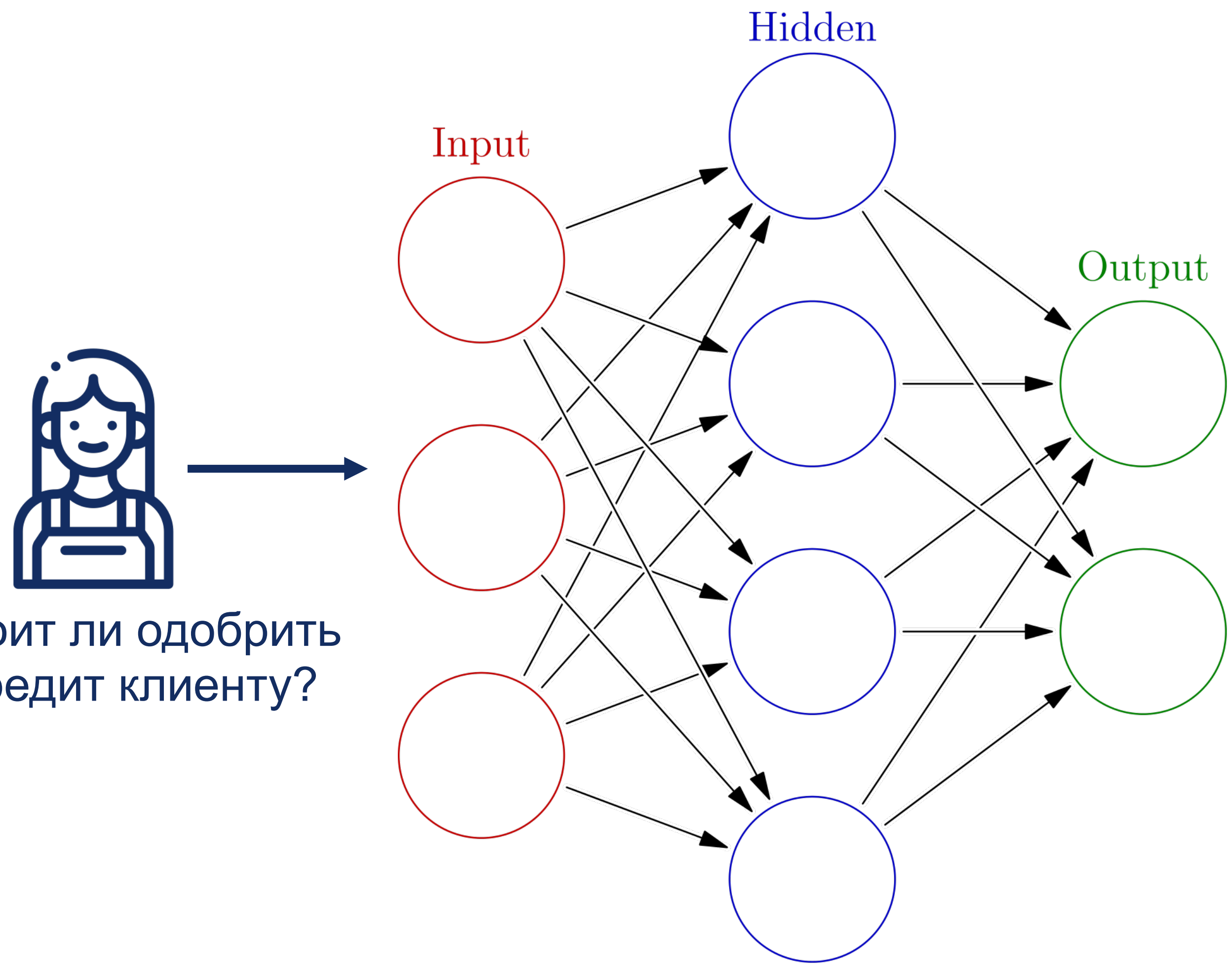
Обучение с учителем

Нейронные сети

Стоит ли одобрить
Кредит клиенту?

Нейронные сети

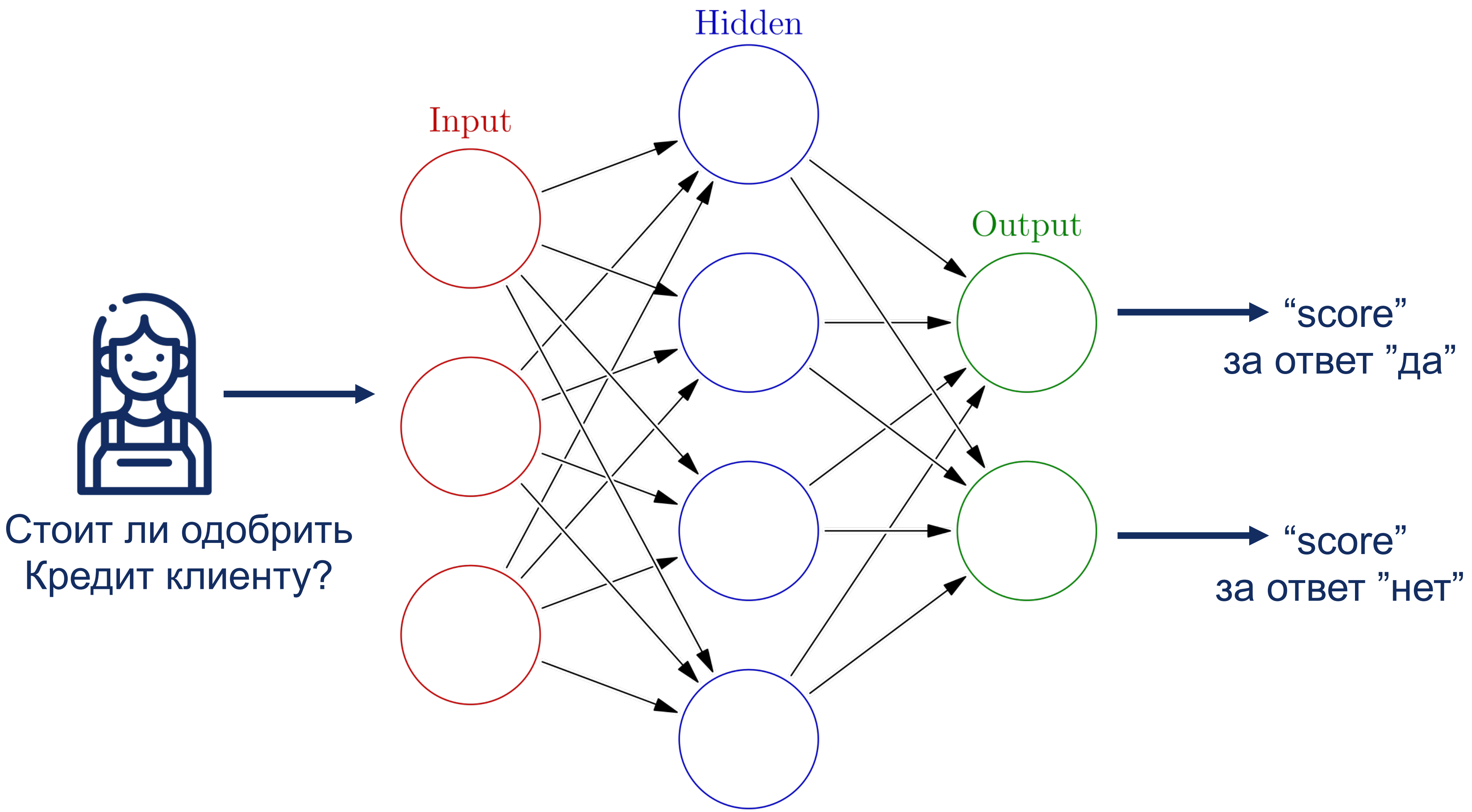
Обучение с учителем



Стоит ли одобрить Кредит клиенту?

Обучение с учителем

Нейронные сети



Линейные модели: построение

To take away:

- Алгоритмов очень много, области применения сильно различаются
- Все алгоритмы основаны на принципах базовой логики и математике
- Практически для всех алгоритмов есть готовая реализация (часто таких реализаций много), этим нужно пользоваться!

Линейные модели: интуиция

Линейные модели: интуиция

? Давайте обучим модель?

Пример: занимать ли очередь в банке?

Признаки (1/0)

Вы свободны в данный момент

Вы голодны

Вам хочется спать

В очереди меньше 3х человек

Пример: занимать ли очередь в банке?

Признаки (1/0)

+1

Вы свободны в данный момент

Вы голодны

Вам хочется спать

В очереди меньше 3х человек

Пример: занимать ли очередь в банке?

Признаки (1/0)

+1

Вы свободны в данный момент

-2

Вы голодны

Вам хочется спать

В очереди меньше 3х человек

Пример: занимать ли очередь в банке?

Признаки (1/0)

+1

Вы свободны в данный момент

-2

Вы голодны

-3

Вам хочется спать

В очереди меньше 3х человек

Пример: занимать ли очередь в банке?

Признаки (1/0)

+1

Вы свободны в данный момент

-2

Вы голодны

-3

Вам хочется спать

+4

В очереди меньше 3х человек

Пример: занимать ли очередь в банке?

Линейные модели

Порог для решающего правила: 0
Если сумма больше 0 – занимаем очередь!

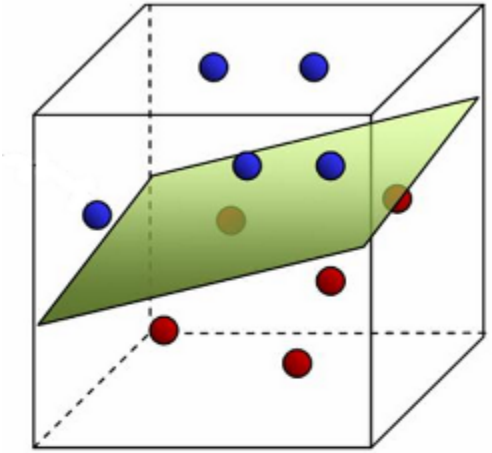
Линейные модели: интуиция

Линейные модели

$$a(x) = \begin{cases} 1, \text{если } f(x) \geq 0 \\ -1, \text{если } f(x) < 0 \end{cases}$$

$$\mathrm{f(x) = w_0 + w_1 x_1 + \cdots + w_n x_n}$$

Геометрическая интерпретация: разделяем классы плоскостью



Пример: скоринговые карты

Показатель	Диапазон значений
Возраст заёмщика	До 35 лет
	От 35 до 45 лет
	От 45 и старше
Образование	Высшее
	Среднее специальное
	Среднее
Состоит ли в браке	Да
	Нет
Наличие кредита в прошлом	Да
	Нет
Стаж работы	До 1 года
	От 1 до 3 лет
	От 3 до 6 лет
	Свыше 6 лет
Наличие автомобиля	Да
	Нет

Пример: скоринговые карты

Показатель	Диапазон значений	Скоринг-балл
Возраст заёмщика	До 35 лет	7,60
	От 35 до 45 лет	29,68
	От 45 и старше	15,87
Образование	Высшее	29,82
	Среднее специальное	20,85
	Среднее	22,71
Состоит ли в браке	Да	29,46
	Нет	9,38
Наличие кредита в прошлом	Да	40,55
	Нет	13,91
Стаж работы	До 1 года	15,00
	От 1 до 3 лет	18,14
	От 3 до 6 лет	19,85
	Свыше 6 лет	23,74
Наличие автомобиля	Да	51,69
	Нет	15,93

Пример: скоринговые карты

Почему нельзя продолжать так же?

- Сложно настраивать вручную
- Требуется эксперт в области
- Требуется проверка на данных и уточнение весов (эксперт может что-то не учесть)

Пример: скоринговые карты

Почему нельзя продолжать так же?

- Сложно настраивать вручную
- Требуется эксперт в области
- Требуется проверка на данных и уточнение весов (эксперт может что-то не учесть)

Решение

Автоматизируем подбор параметров: придумаем функцию от параметров, которую надо минимизировать, и используем методы численной оптимизации

Пример: скоринговые карты

Почему нельзя продолжать так же?

- Сложно настраивать вручную
- Требуется эксперт в области
- Требуется проверка на данных и уточнение весов (эксперт может что-то не учесть)

Решение

Автоматизируем подбор параметров: придумаем функцию от параметров – Q(a), которую надо минимизировать Q(a)->min, и используем методы численной оптимизации

Линейные модели: построение

Линейные модели: построение

Линейная регрессия

$$a(x) = \langle w, x \rangle + w_0$$

Линейные модели: построение

Линейная регрессия

$$a(x) = \langle w, x \rangle + w_0$$

А как получить ответ в задаче классификации?

Линейные модели: построение

Линейная классификация

$$a(x) = \langle w, x \rangle + w_0$$

А как получить ответ в задаче классификации?

$$a(x) = sign(\langle w, x \rangle + w_0)$$

Если скалярное произведение неотрицательное – класс 1, в противном случае класс 0

Линейные модели: построение

Линейная модель

$$a(x) = \langle w, x \rangle + w_0$$

$$a(x) = sign(\langle w, x \rangle + w_0)$$

Вид модели задан. Теперь нужно выписать задачу оптимизации для настройки параметров модели (весов).

Линейные модели: построение

Отступ

Отступом алгоритма $a(x) = sign\{f(x)\}$ на объекте x_i называется величина

$$M_i = y_i f(x_i)$$

(y_i - класс, к которому относится x_i)

$$M_i \leq 0 \Leftrightarrow y_i \neq a(x_i)$$
$$M_i > 0 \Leftrightarrow y_i = a(x_i)$$

Линейные модели: построение

Эмпирический риск

$$Q(w) = \sum_{i=1}^{\ell} [M_i(w) < 0]$$

Линейные модели: построение

Функция потерь

$$Q(w) = \sum_{i=1}^{\ell} [M_i(w) < 0] \leqslant \tilde{Q}(w) = \sum_{i=1}^{\ell} \mathscr{L}(M_i(w)) \to \min_{w};$$

Функция эмпирического риска

Функция потерь

Линейные модели: построение

Функция потерь

$$Q(w) = \sum_{i=1}^{\ell} [M_i(w) < 0] \leqslant \tilde{Q}(w) = \sum_{i=1}^{\ell} \mathscr{L}(M_i(w)) \to \min_w;$$

Общий вид задачи оптимизации задан, остается несколько степеней свободы:

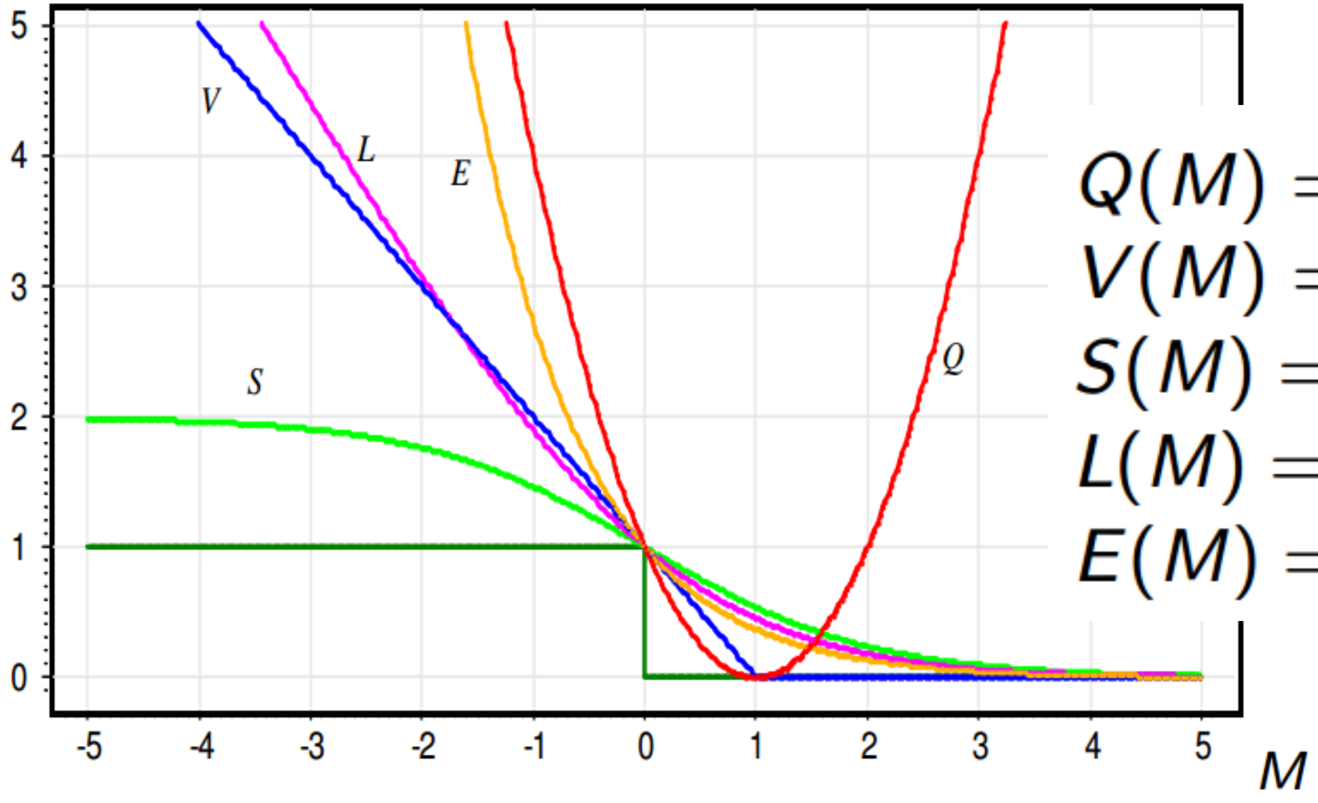
- функция потерь (L)
- дополнительные ограничения
- метод оптимизации

Давайте обсудим варианты.

Функция потерь

Линейные модели: построение

$$Q(w) = \sum_{i=1}^{\ell} [M_i(w) < 0] \leqslant \tilde{Q}(w) = \sum_{i=1}^{\ell} \mathscr{L}(M_i(w)) \to \min_w;$$



$$Q(M) = (1 - M)^2$$
$$V(M) = (1 - M)_+$$
$$S(M) = 2(1 + e^M)^{-1}$$
$$L(M) = \log_2(1 + e^{-M})$$
$$E(M) = e^{-M}$$

Линейные модели: построение

Задача оптимизации

$$a(x) = \langle w, x \rangle + w_0$$

$$Q = \sum_{i=1}^{l} L\big(y_i, a(x_i)\big) \to \min_{w}$$

Линейные модели: построение

Задача оптимизации

$$a(x) = \langle w, x \rangle + w_0$$

$$Q = \sum_{i=1}^{l} L\big(y_i, a(x_i)\big) \to \min_{w}$$

$$L\big(y_i, a(x_i)\big) = (y_i - a(x_i))^2$$

$$L\big(y_i, a(x_i)\big) = |y_i - a(x_i)|$$

Линейные модели: построение

Задача оптимизации

$$a(x) = \langle w, x \rangle + w_0$$

$$Q = \sum_{i=1}^{l} L\big(y_i, a(x_i)\big) \to \min_{w}$$

$$L\big(y_i, a(x_i)\big) = \ (y_i - a(x_i))^2$$

$$L\big(y_i, a(x_i)\big) = |y_i - a(x_i)|$$

$$V(w) = \ \|w\|_{l2}^2 = \sum_{n=1}^{d} w_n^2$$

$$V(w) = \|w\|_{l1} = \sum_{n=1}^{d} |w_n|$$

Линейные модели: построение

Задача оптимизации

$$a(x) = \langle w, x \rangle + w_0$$

$$Q = \sum_{i=1}^{l} L\big(y_i, a(x_i)\big) + \gamma V(w) \to \min_{w}$$

Гребневая регрессия (Ridge regression):

$$V(w) = \|w\|_{l2}^2 = \sum_{n=1}^{d} w_n^2$$

LASSO (least absolute shrinkage and selection operator):

$$V(w) = \|w\|_{l1} = \sum_{n=1}^{d} |w_n|$$

Линейные модели: построение

Задача оптимизации

$$a(x) = \langle w, x \rangle + w_0$$

$$Q = \sum_{i=1}^{l} L\big(y_i, a(x_i)\big) \to \min_{w}$$

$$L\big(y_i, a(x_i)\big) = \left(y_i - a(x_i)\right)^2$$

А без регуляризатора и с квадратичными потерями получаем привычную нам линейную регрессию

Линейные модели: построение

Конструирование линейных моделей

Классификатор	Функция потерь	Регуляризатор
SVM (Support vector machine, метод опорных векторов)	$L(M) = \max\{0, 1 - M\} = = (1 - M)_+$	$\sum_{k=1}^{m} {w_k}^2$
Логистическая регрессия	$L(M) = \log(1 + e^{-M})$	Обычно $\sum_{k=1}^{m} {w_k}^2$ или $\sum_{k=1}^{m} \lvert w_k \rvert$

Линейные модели: построение

Решение задачи оптимизации

Модель: $y_i \approx \hat{y}_i = < w, x_i > + w_0$

Линейные модели: построение

Решение задачи оптимизации

Модель: $y_i \approx \hat{y}_i =< w, x_i > +w_0$

Если добавить $x_{i0} = 1$:

Линейные модели: построение

Решение задачи оптимизации

Модель: $y_i \approx \hat{y}_i = < w, x_i > + w_0$

Если добавить $x_{i0} = 1$:

$y_i \approx \hat{y}_i = < w, x_i >$

Линейные модели: построение

Решение задачи оптимизации

Модель: $y_i \approx \hat{y}_i =< w, x_i > +w_0$

Если добавить $x_{i0} = 1$:

$y_i \approx \hat{y}_i =< w, x_i >$

$$y_1 \approx \hat{y}_1 = {x_1}^T w$$
$$\dots$$
$$y_i \approx \hat{y}_i = {x_i}^T w$$
$$\dots$$
$$y_l \approx \hat{y}_l = {x_l}^T w$$

Линейные модели: построение

Решение задачи оптимизации

Матричная запись

$$\begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_l \end{pmatrix} \approx \begin{pmatrix} \widehat{y_1} \\ \widehat{y_2} \\ \dots \\ \widehat{y_l} \end{pmatrix} = \begin{pmatrix} {x_1}^T \\ {x_2}^T \\ \dots \\ {x_l}^T \end{pmatrix} w$$

$$y \approx \hat{y} = Fw$$

$$w = \underset{w}{\operatorname{argmin}} ||y - \hat{y}||^2$$

Линейные модели: построение

Решение задачи оптимизации

$$\frac{\partial(Fw - y)^2}{\partial w} = 2F^T(Fw - y) = 0$$

$$F^T F w = F^T y$$

$$\boxed{w = (F^T F)^{-1} F^T y}$$

Линейные модели: построение

Решение задачи оптимизации

Если добавить ℓ_2 регуляризацию

$$\frac{\partial (Fw - y)^2 + \gamma w^2}{\partial w} = 2F^T(Fw - y) + 2\gamma w = 0$$

$$(F^T F + \gamma I)w = F^T y$$

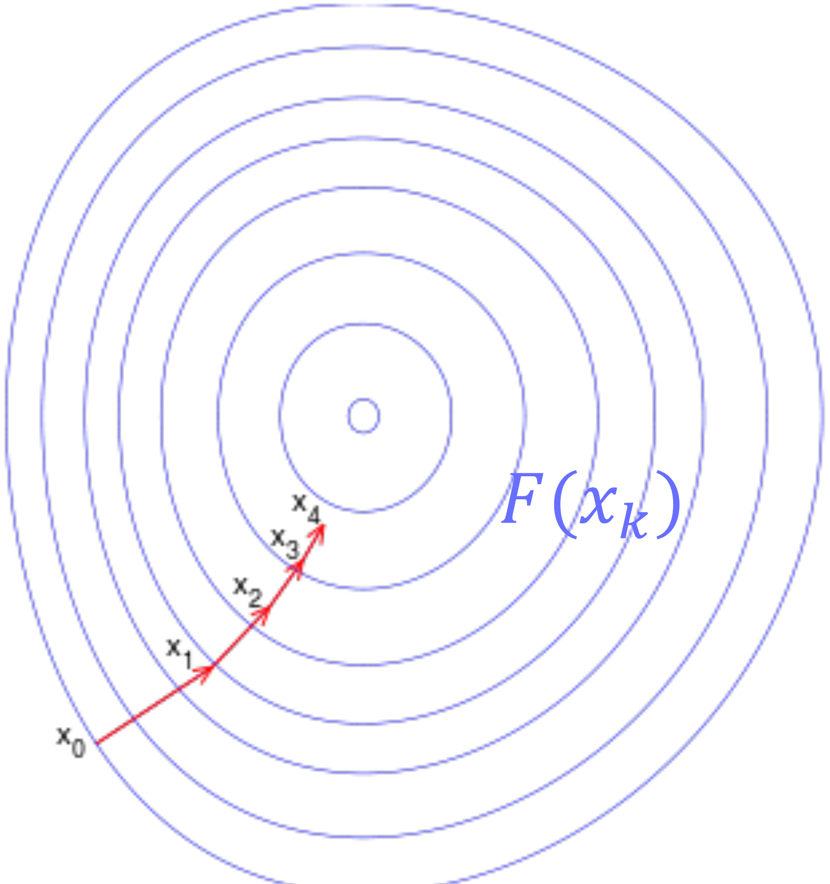
$$\boxed{w = (F^T F + \gamma I)^{-1} F^T y}$$

Линейные модели: построение

Решение задачи оптимизации

Численное решение возможно с помощью градиентного спуска (GD, Gradient Decent)

$$x_{k+1} = x_k - \gamma_k \nabla F(x_k)$$

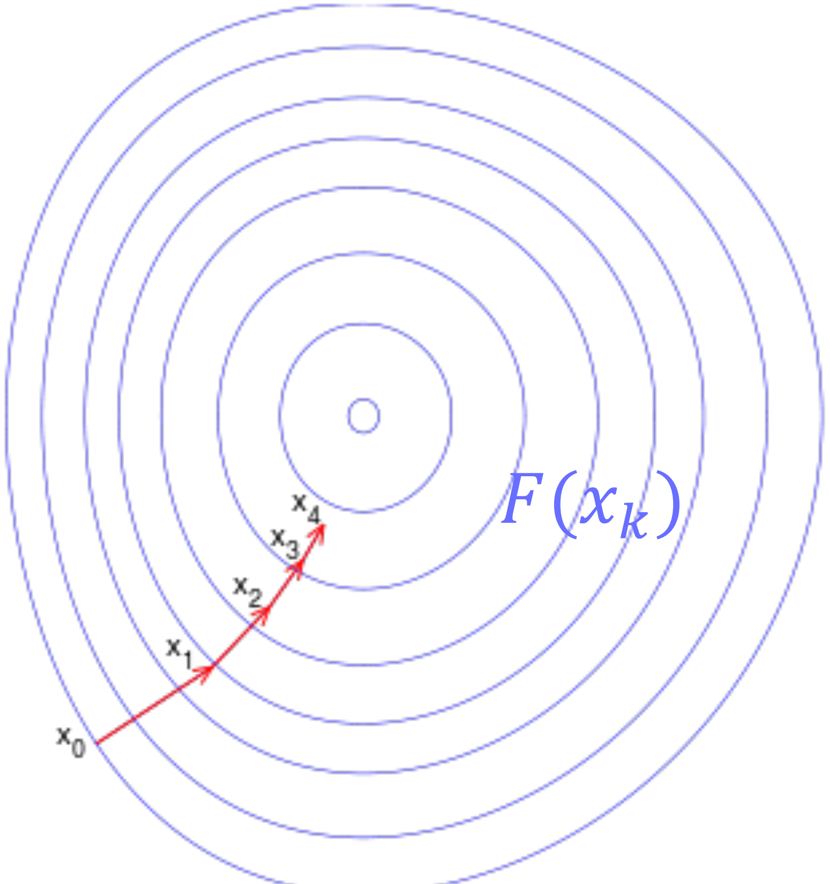


Линейные модели: построение

Градиентный спуск (GD)

$$\nabla_w \tilde{Q} = \sum_{i=1}^{l} \nabla L(M_i) = \sum_{i=1}^{l} L'(M_i) \frac{\partial M_i}{\partial w}$$

$$x_{k+1} = x_k - \gamma_k \nabla F(x_k)$$



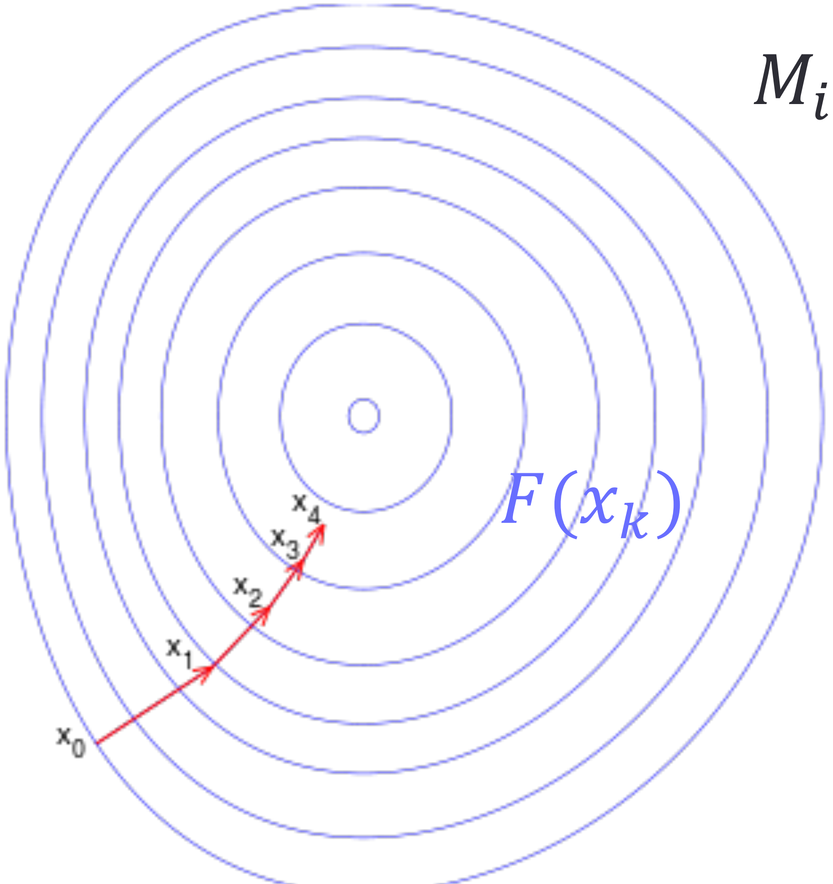
Линейные модели: построение

Градиентный спуск (GD)

$$\nabla_w \tilde{Q} = \sum_{i=1}^{l} \nabla L(M_i) = \sum_{i=1}^{l} L'(M_i) \frac{\partial M_i}{\partial w}$$

$$x_{k+1} = x_k - \gamma_k \nabla F(x_k)$$

$$M_i = y_i \langle w, x_i \rangle \implies \frac{\partial M_i}{\partial w} = y_i x_i$$

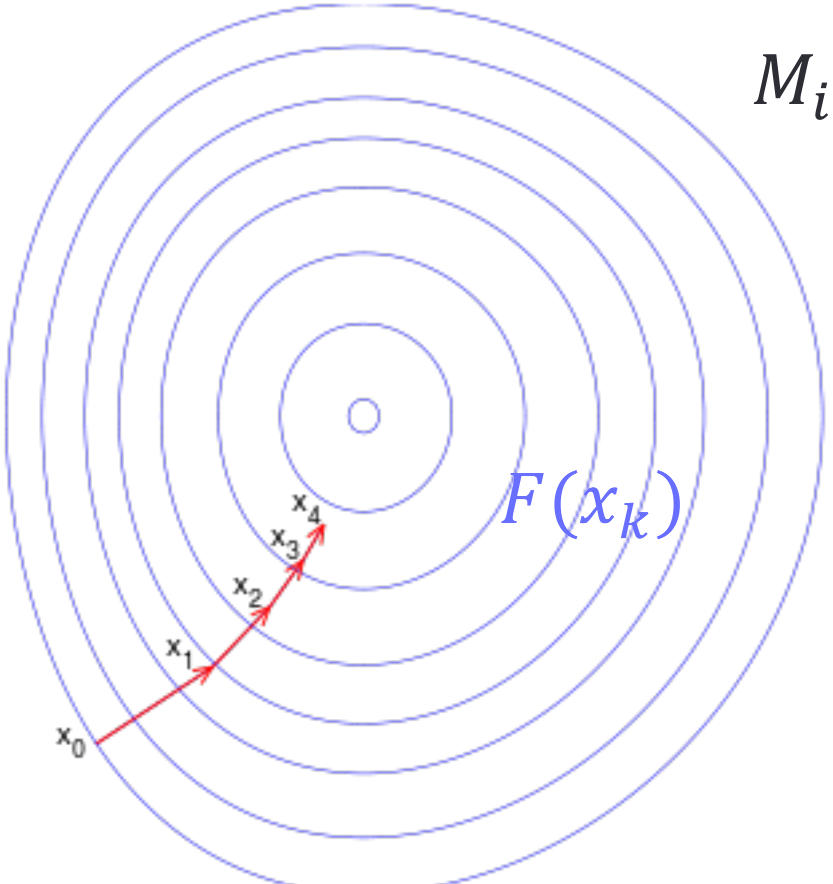


Линейные модели: построение

Градиентный спуск (GD)

$$\nabla_w \tilde{Q} = \sum_{i=1}^{l} \nabla L(M_i) = \sum_{i=1}^{l} L'(M_i) \frac{\partial M_i}{\partial w}$$

$$x_{k+1} = x_k - \gamma_k \nabla F(x_k)$$



$$M_i = y_i \langle w, x_i \rangle \Longrightarrow \frac{\partial M_i}{\partial w} = y_i x_i$$

$$\nabla \tilde{Q} = \sum_{i=1}^{l} y_i x_i L'(M_i)$$

Линейные модели: построение

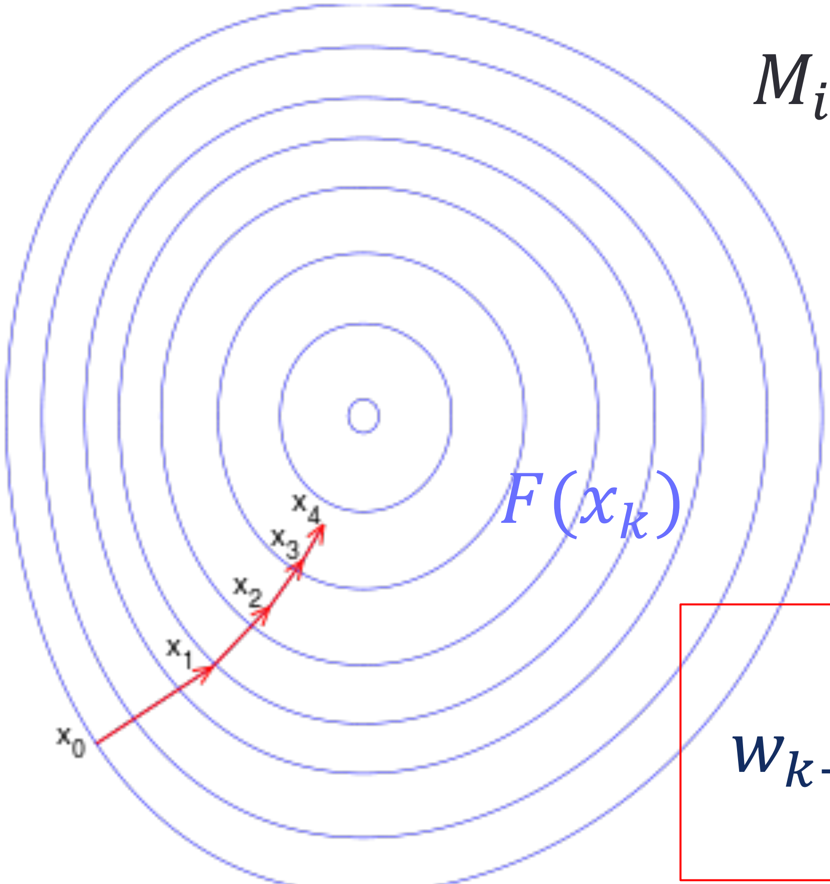
Градиентный спуск (GD)

$$\nabla_w \tilde{Q} = \sum_{i=1}^{l} \nabla L(M_i) = \sum_{i=1}^{l} L'(M_i)\frac{\partial M_i}{\partial w}$$

$$x_{k+1} = x_k - \gamma_k \nabla F(x_k)$$

$$M_i = y_i\langle w, x_i\rangle \Longrightarrow \frac{\partial M_i}{\partial w} = y_i x_i$$

$$\nabla \tilde{Q} = \sum_{i=1}^{l} y_i x_i L'(M_i)$$



$$\boxed{w_{k+1} = w_k - \gamma_k \sum_{i=1}^{l} y_i x_i L'(M_i)}$$

Линейные модели: построение

Стохастический градиент (SGD)

$$w_{k+1} = w_k - \gamma_k \cancel{\sum_{i=1}^{l}} y_i x_i L'(M_i)$$

$$\boxed{w_{k+1} = w_k - \gamma_k y_i x_i L'(M_i)}$$

x_i – случайный элемент обучающей выборки

Линейные модели: построение

To take away:

Линейные модели имеют ряд преимуществ:

- Простота реализации
- Скорость работы
- Хорошее качество, когда много признаков
- Приемлемое качество, когда мало данных

Особенности применения

- Модель может оказаться слишком простой для задачи
- Требуется бороться с переобучением: регуляризация, масштабирование признаков

Машинное обучение: линейные модели классификации и регрессии

Спасибо!

Эмели Драль