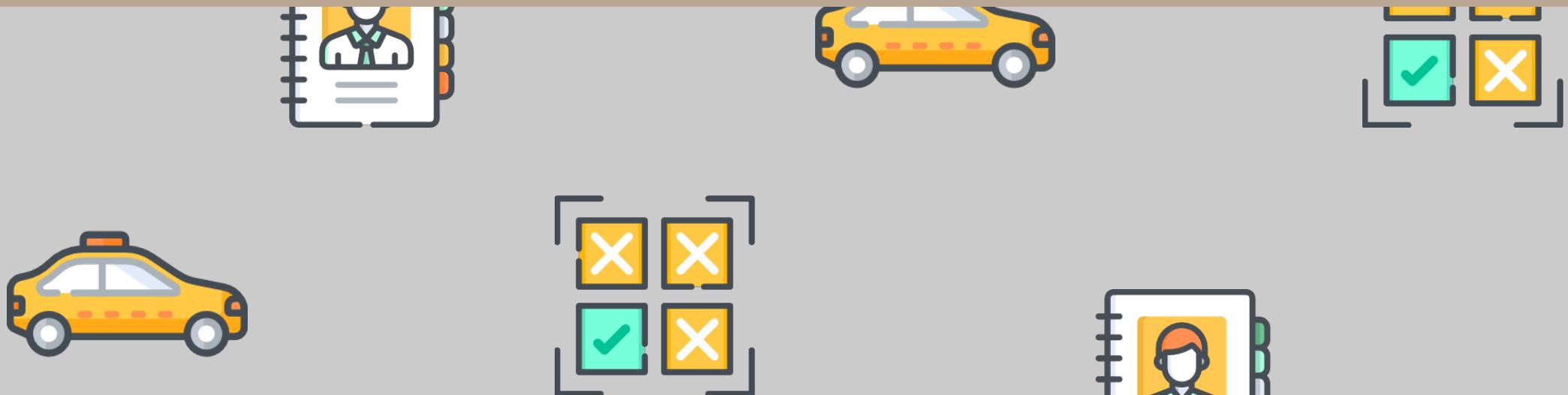


Нейронные сети на практике



0. Зачем мне эта лекция?



Посыл

- Многие задачи ML в индустрии решаются с помощью методов DL
- Очень много теоретической информации про сети
- Неясно, что нужно знать и уметь, чтобы быть крутым дата саентистом

Что нужно зафиксировать в конце лекции?

- Минимум теории про сети, перечень нужных понятий
- Какие задачи в индустрии решают с помощью DL
- Что нужно знать, чтобы успешно их решать
- Где возникают нюансы в применении методов
- Зачем и где саморазвиваться?

1. Какие задачи про DL бывают в индустрии?



Тексты



background
Approach
semantic
search
practical

- Поддержка
- Чаты между пользователями
- Анализ медиа (что про нашу компанию пишут в сети)



Картинки

- Фотки машин
- Биометрия
- Фотки блюд



Акции

The image shows a mobile application interface for food delivery. At the top, there are two promotional banners for different restaurants:

- PinzaUno**: Offers 15–25 minutes delivery. The banner features several plates of Italian pasta like spaghetti and fettuccine, garnished with basil, cheese, and tomatoes. It includes a star rating of 4.2 and a price starting from 0 rubles.
- Djagannat**: Offers 15–25 minutes delivery. The banner features plates of various dishes, including what looks like a salad and some stir-fried vegetables. It includes a star rating of 4.2 and a price starting from 0 rubles.

Below the banners, there is a section for Italian cuisine (Итальянская) featuring pizza (Пицца) with a price of 222 rubles.

Видео



- То же, что и картинки,
только больше
информации

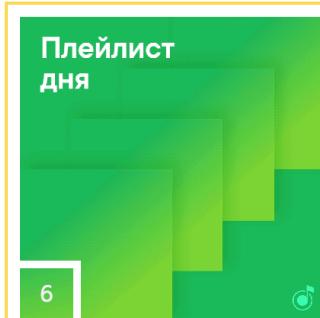
Аудио



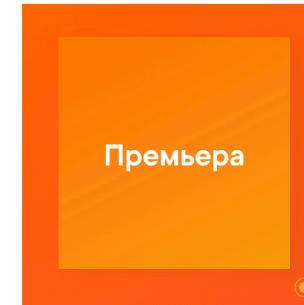
- Звонки в коллцентры
- Музыка в плейлисте



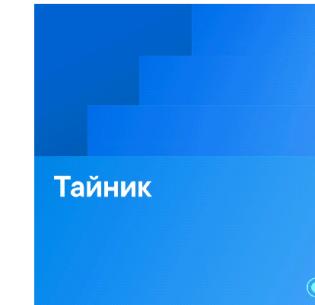
Умные плейлисты для вас, Татьяна



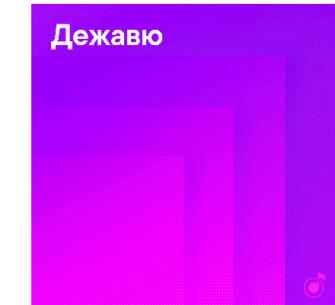
Плейлист дня
Каждый день — новый. Каждый день — ваш!
Обновлён сегодня



Премьера •
Только новинки, подобранные по вашим предпочтениям
Обновлён сегодня



Тайник
Треки из вашей фонотеки, которые вы ещё не послушали
Обновлён во вторник



Дежавю •
Вы ещё не слушали эти треки, но, похоже, вам они понравятся
Обновлён в понедельник

2. Линейные модели и нейронные сети

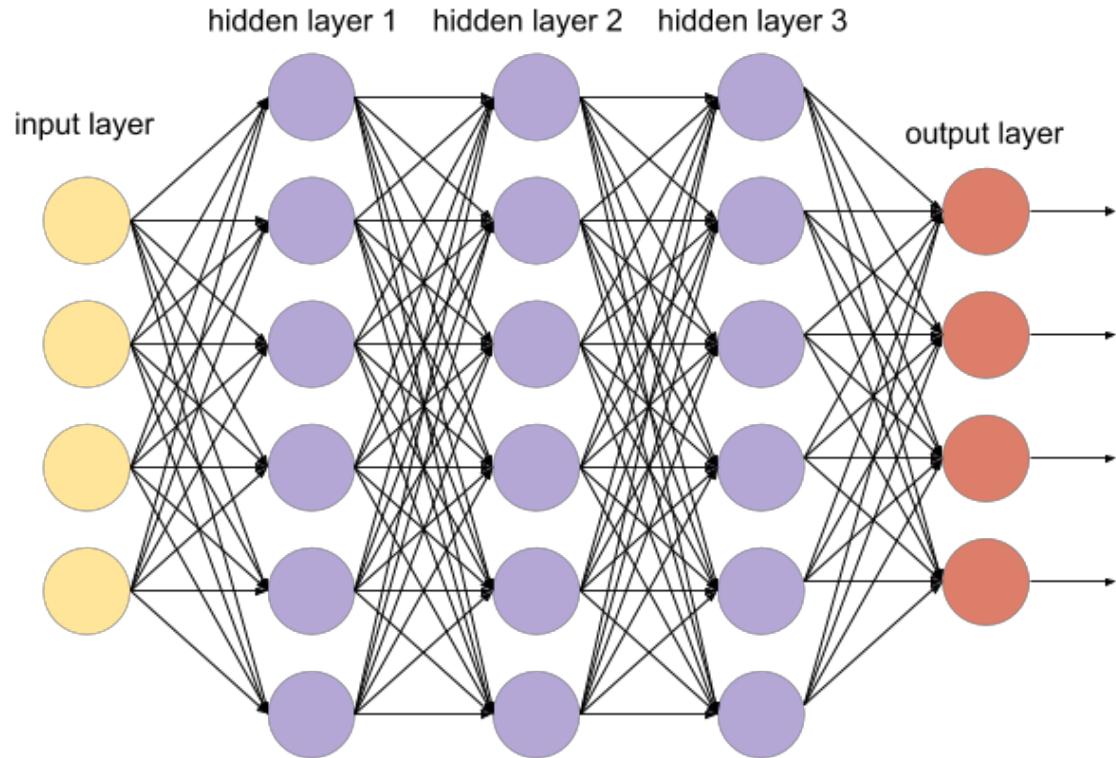


Линейная модель

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

$$Q = \sum_{i=1}^l L(y_i, a(x_i)) \rightarrow \min_w$$

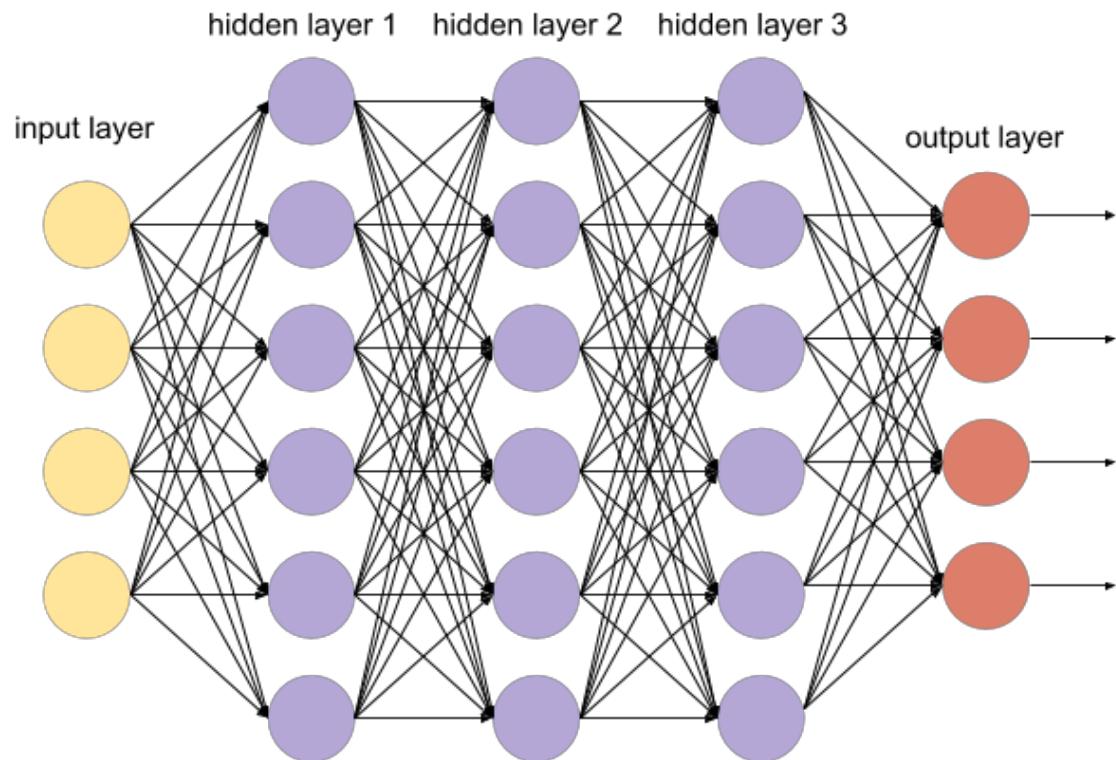
Нейронная сеть



$$h_t = f(Wh_{t-1} + b)$$

$$a(x) = f_3(W_3f_2(W_2f_1(W_1f_0(W_0x + b_0) + b_1) + b_2) + b_3)$$

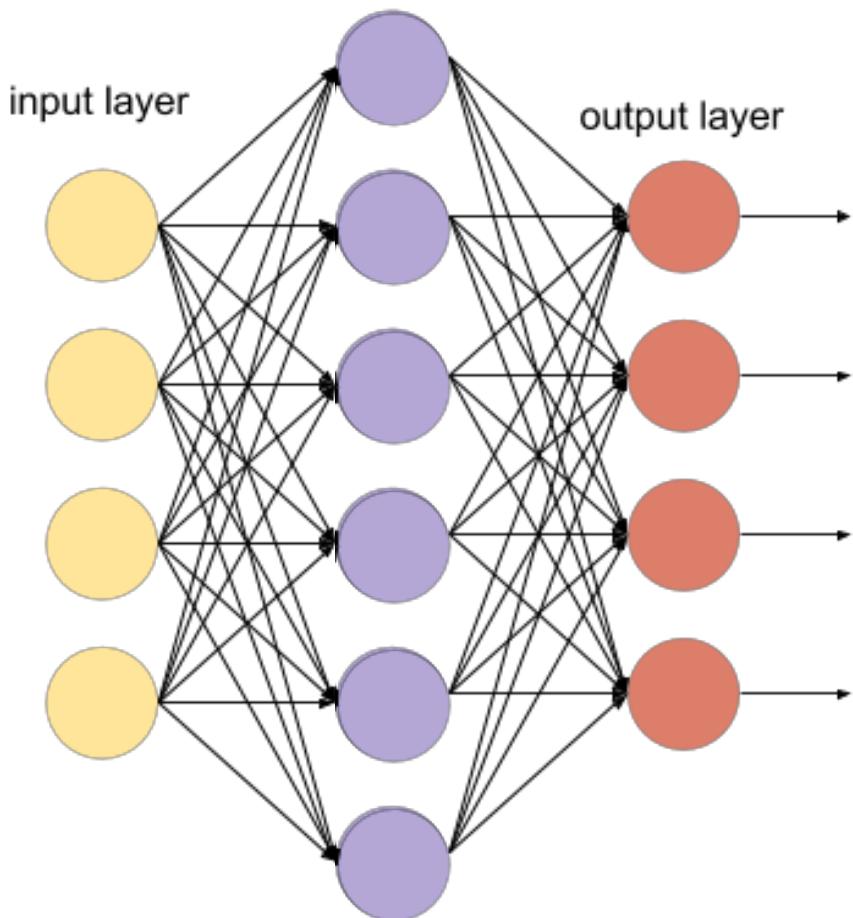
Нейронная сеть



$$h_t = f(Wh_{t-1} + b)$$

$$a(x) = f_3(W_3f_2(W_2f_1(W_1f_0(W_0x + b_0) + b_1) + b_2) + b_3)$$

Универсальная теорема аппроксимации



Капелька оптимизма перед обсуждением обучения:
В 1989 г. Джорджем Цыбенко (George Cybenko) была доказана Universal Approximation Theorem

Нестрого:

Если нам дана функция f и сказано, с какой точностью ее нужно приблизить (какой бы эта точность ни была) – мы всегда справимся с задачей даже однослоиной нейросетью, т.е. сможем подобрать подходящее количество нейронов и веса

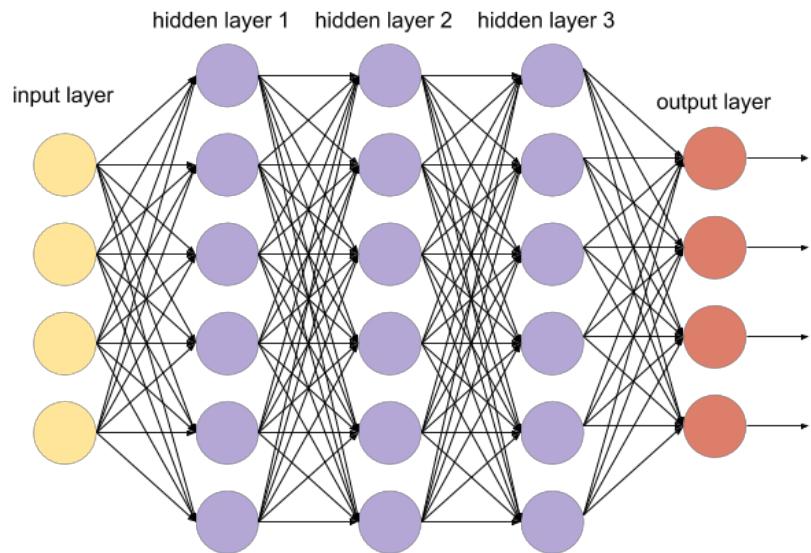
Чуть более строго (для математиков):

f должна быть непрерывна на некотором компакте в R^n и условия теоремы выполняются на нём же

3. Обучение сети

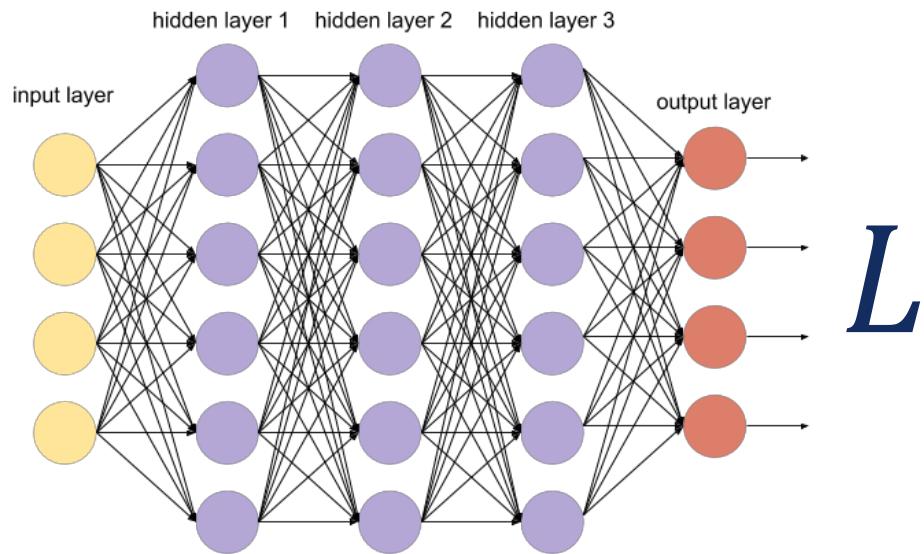


Обучение сети



Задача обучения – настроить веса связей между нейронами на основе обучающей выборки

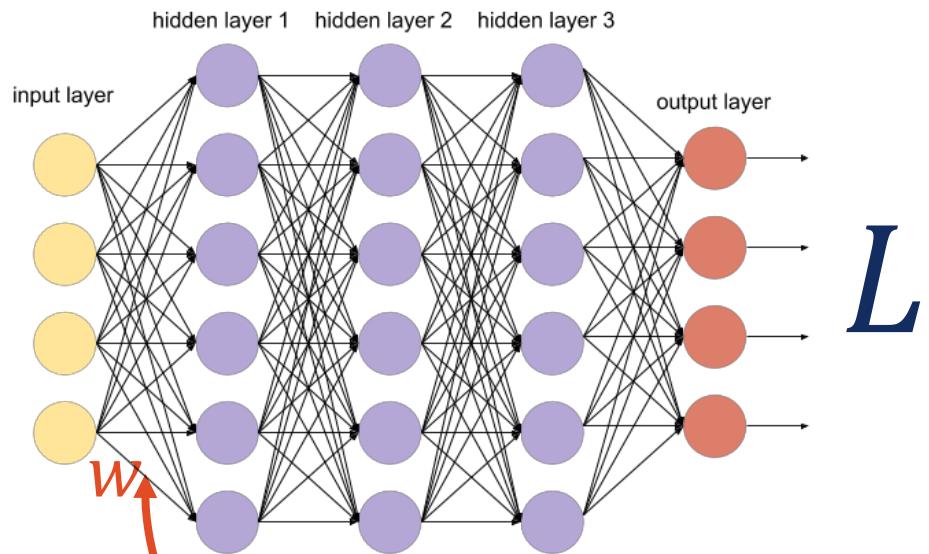
Обучение сети



Задача обучения – настроить веса связей между нейронами на основе обучающей выборки

1. Выбираем функцию потерь

Обучение сети

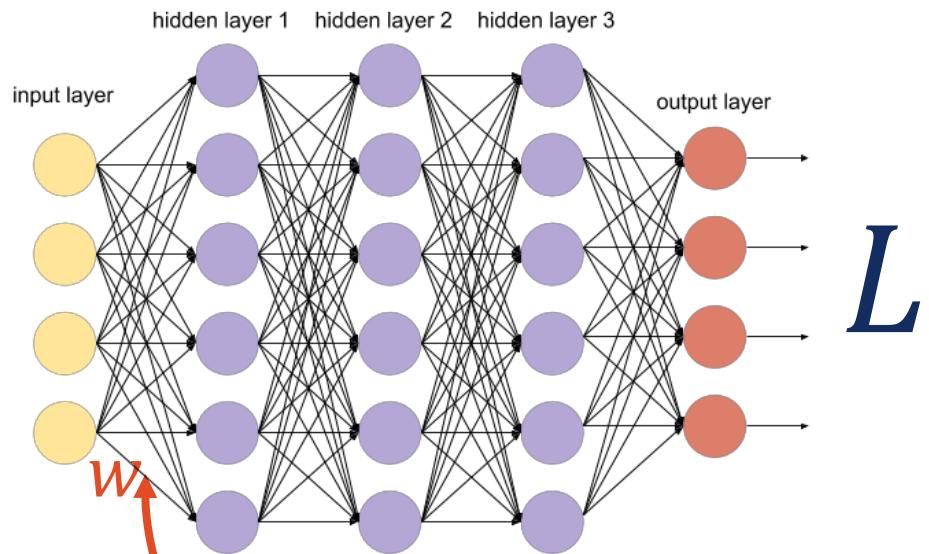


$$w^{(t+1)} = w^{(t)} - \gamma_t \frac{\partial L}{\partial w}$$

Задача обучения – настроить веса связей между нейронами на основе обучающей выборки

1. Выбираем функцию потерь
2. Обучаем веса с помощью SGD

Обучение сети



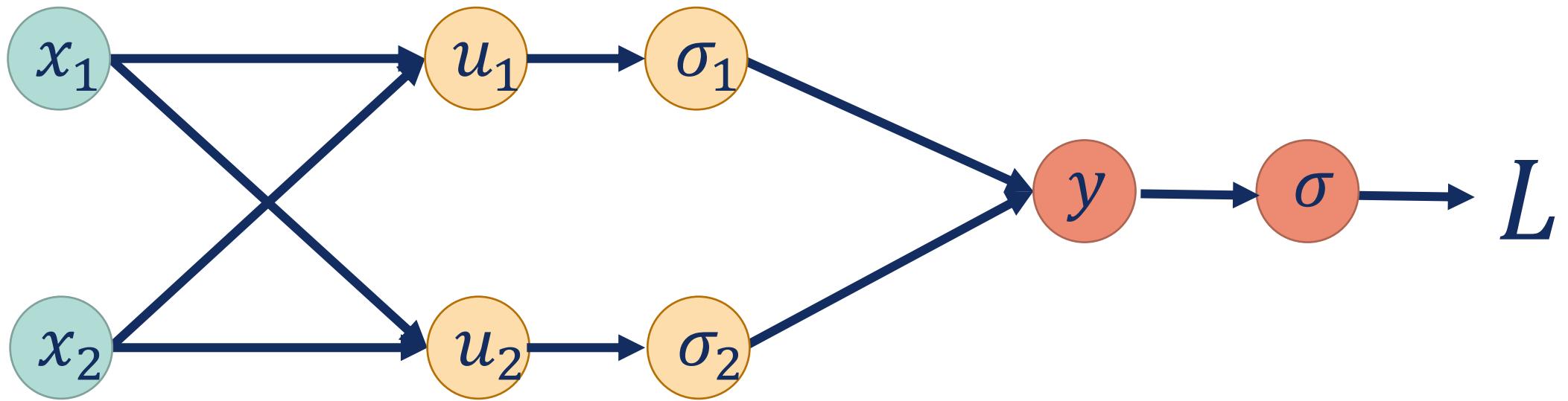
$$w^{(t+1)} = w^{(t)} - \gamma_t \frac{\partial L}{\partial w}$$

Задача обучения – настроить веса связей между нейронами на основе обучающей выборки

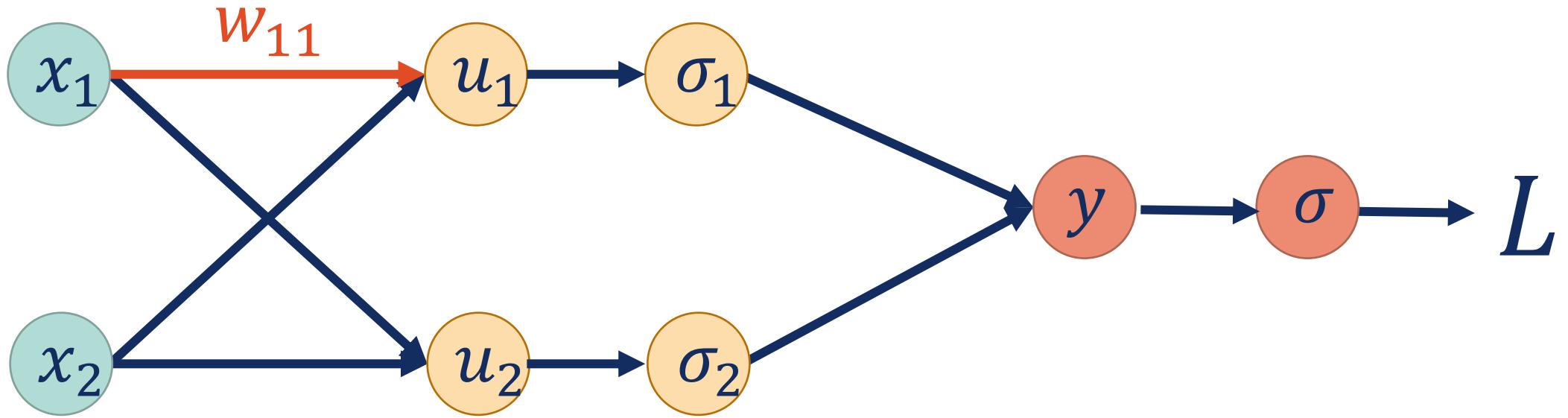
1. Выбираем функцию потерь
2. Обучаем веса с помощью SGD

Проблема: не выписывать же нам все производные аналитически?!

Граф вычислений для нейросети

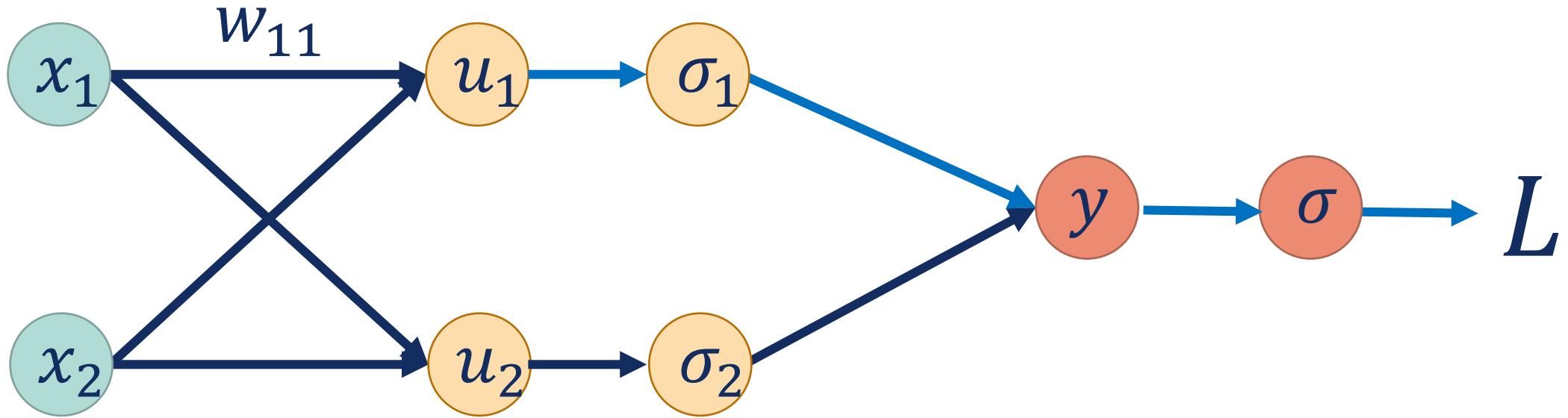


Граф вычислений для нейросети



$$w_{11}^{(t+1)} = w_{11}^{(t)} - \gamma_t \frac{\partial L}{\partial w_{11}}$$

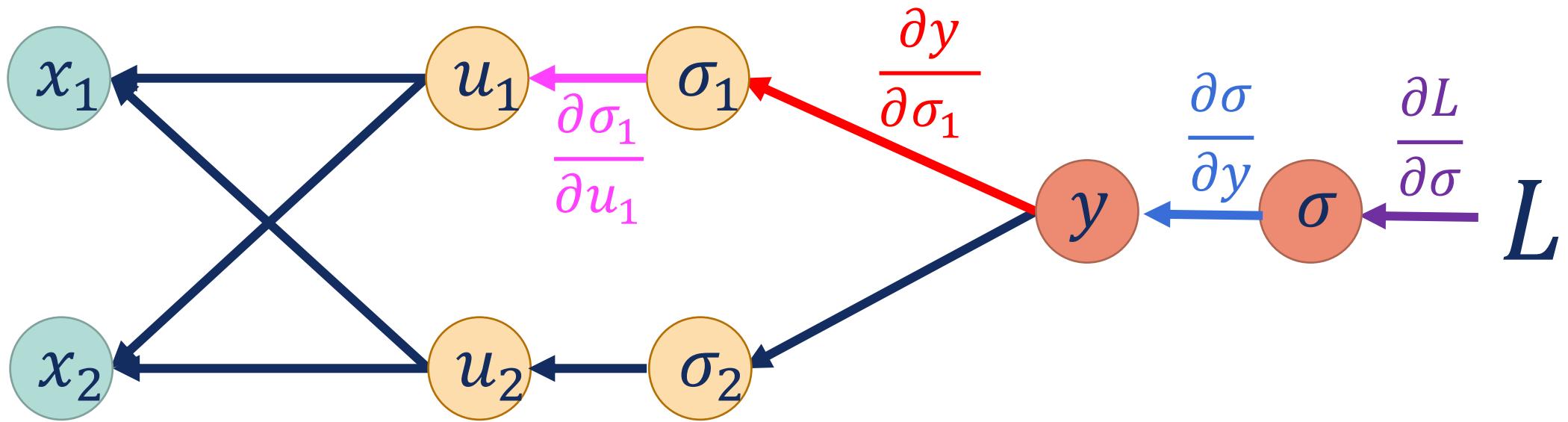
Граф вычислений для нейросети



$$w_{11}^{(t+1)} = w_{11}^{(t)} - \gamma_t \frac{\partial L}{\partial w_{11}}$$

$$\frac{\partial L}{\partial w_{11}} = \frac{\partial L}{\partial u_1} \frac{\partial u_1}{\partial w_{11}} = \frac{\partial L}{\partial u_1} x_1$$

Как вычисляем производную

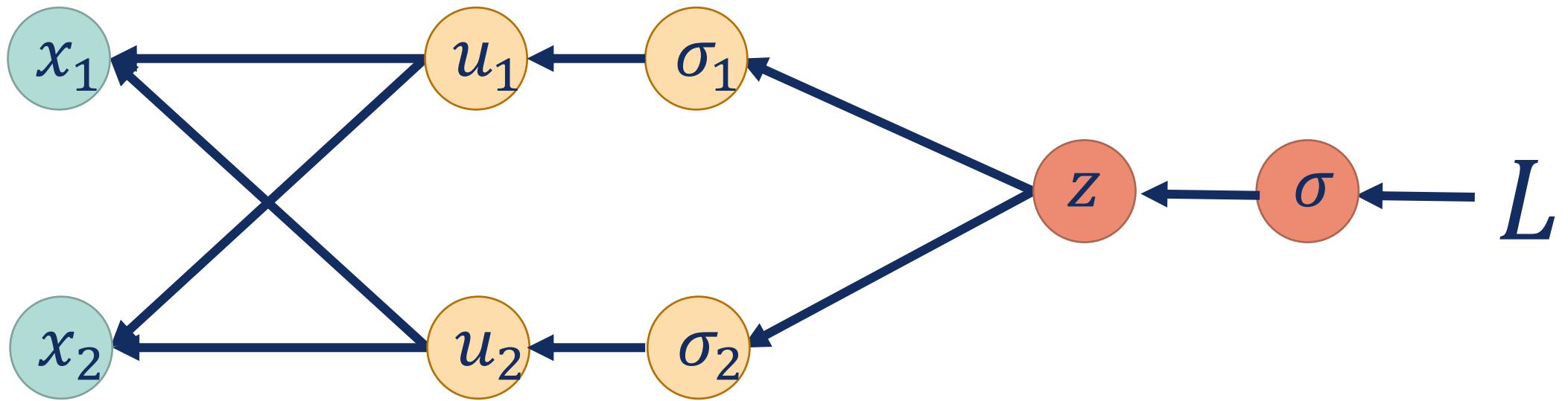


$$\frac{\partial L}{\partial u_1} = \frac{\partial L}{\partial \sigma} \frac{\partial \sigma}{\partial y} \frac{\partial y}{\partial \sigma_1} \frac{\partial \sigma_1}{\partial u_1}$$

Смысл backprop – проходиться по графу с конца и записывать в вершинах графа эти произведения, а не пересчитывать произведение каждый раз заново

Другое название: Error Backpropagation

Раньше этот метод часто называли обратным распространением ошибки

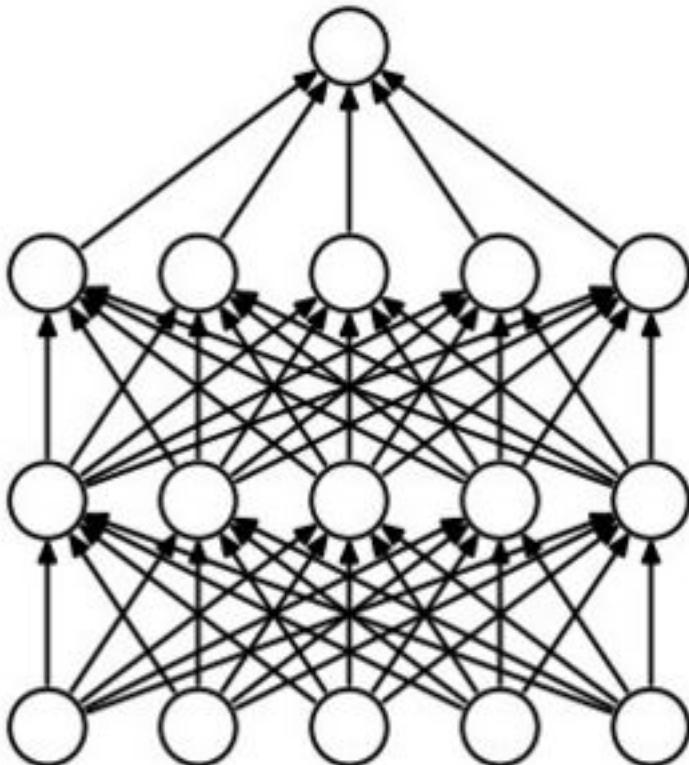


Получается, мы вычисляем ошибку на выходе и «распространяем» ее в обратном направлении (ко входу), вычисляя «ошибки» во всех нейронах по пути

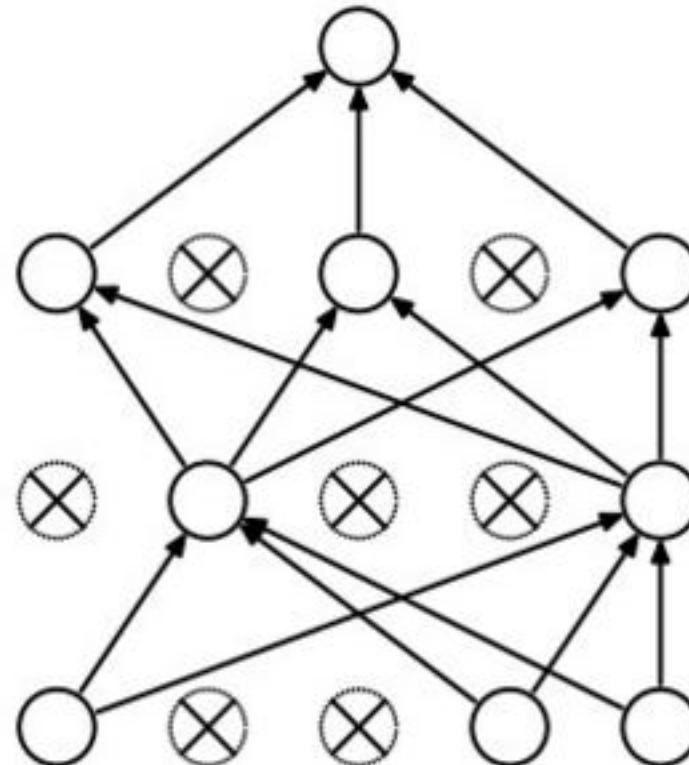
5. Переобучение



Пример регуляризации: dropout



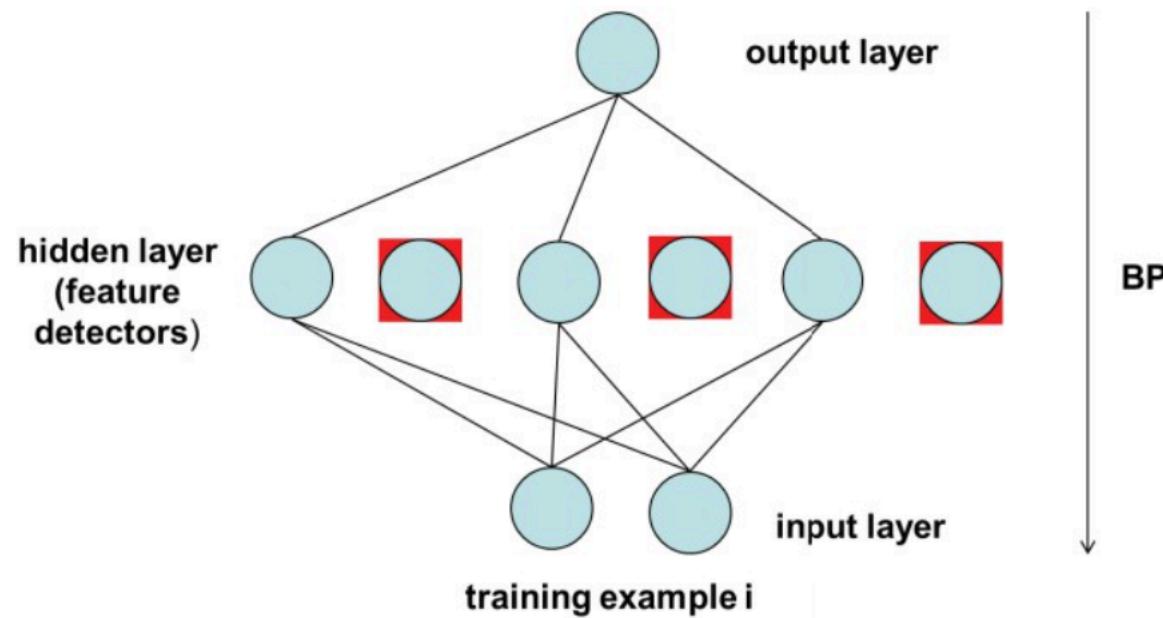
(a) Standard Neural Net



(b) After applying dropout.

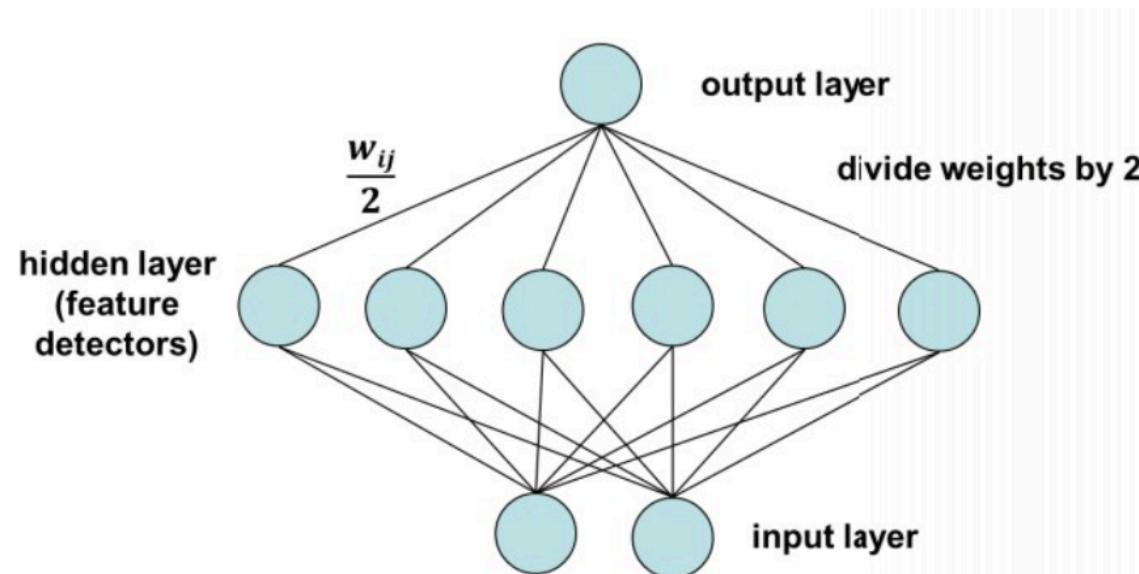
Dropout: обучение

С вероятностью r зануляем
выход каждого нейрона на слое



Dropout: применение

Домножаем выход каждого нейрона на $(1-p)$

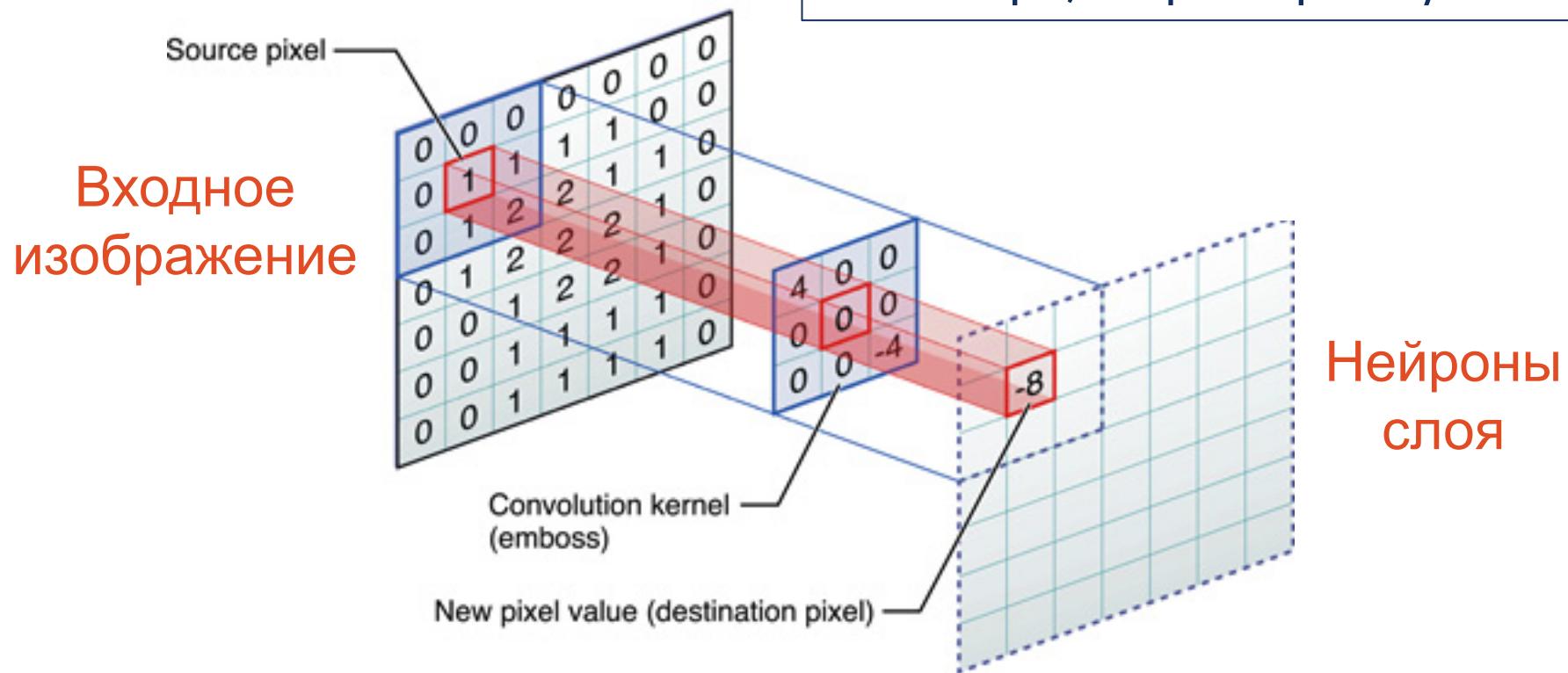


6. Слои - классика



Convolutional

1. Т.к. фильтр – это настраиваемые backprop веса, то **сеть сама «подберет» фильтры**
2. Но т.к. фильтр «замечает» только один паттерн, то фильтров нужно **больше**



Pooling

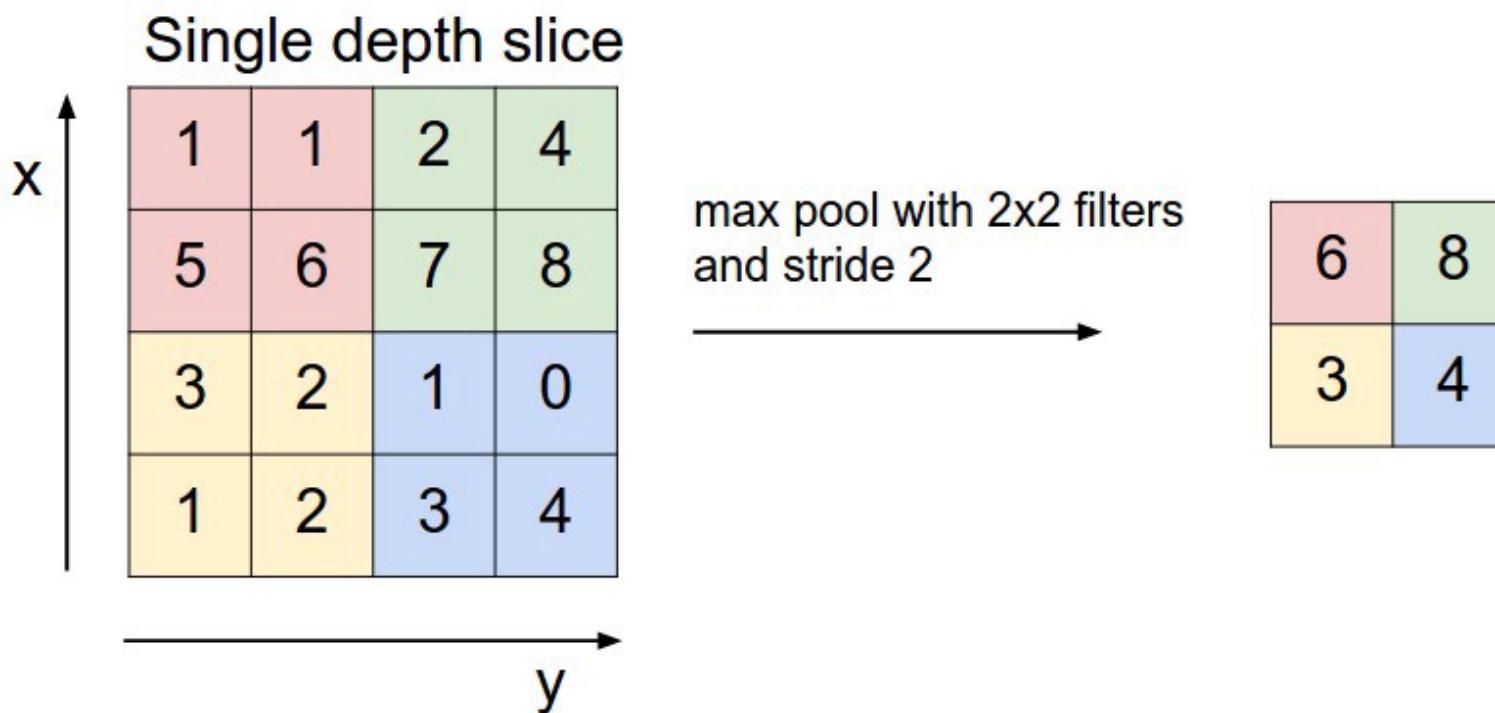
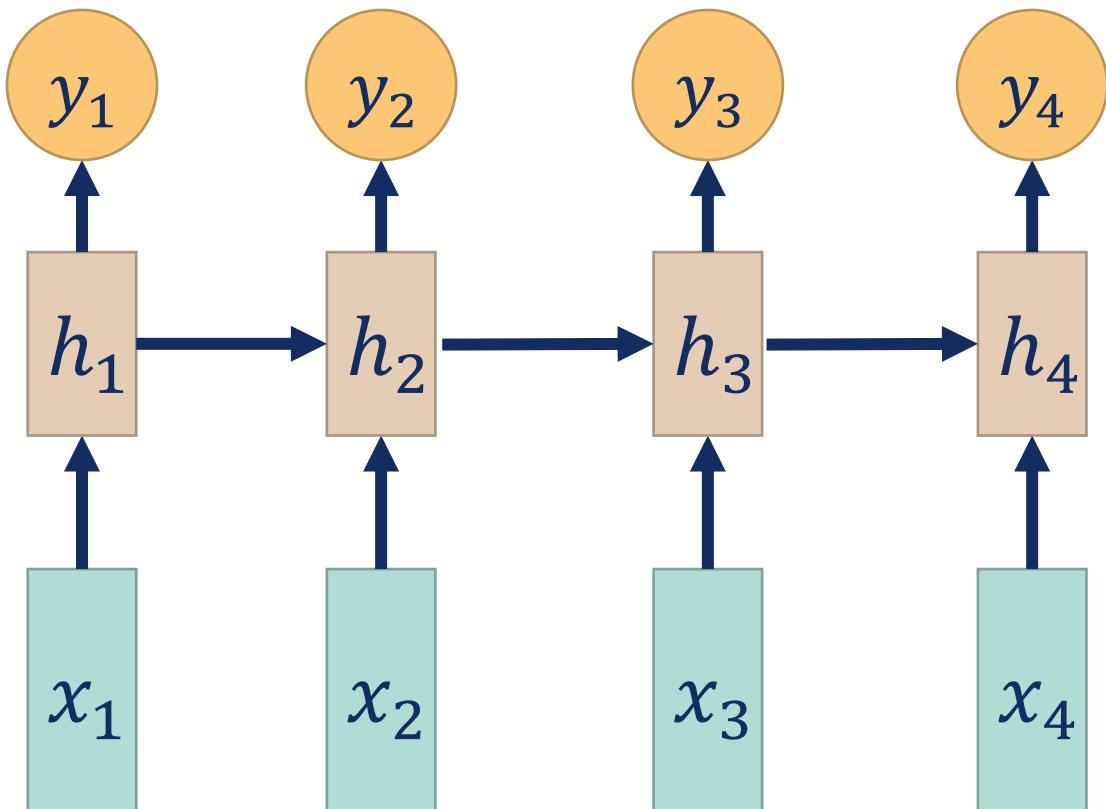
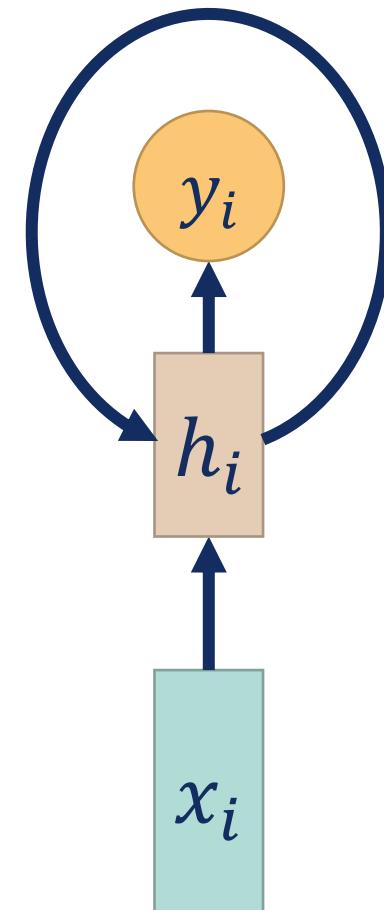


Схема RNN

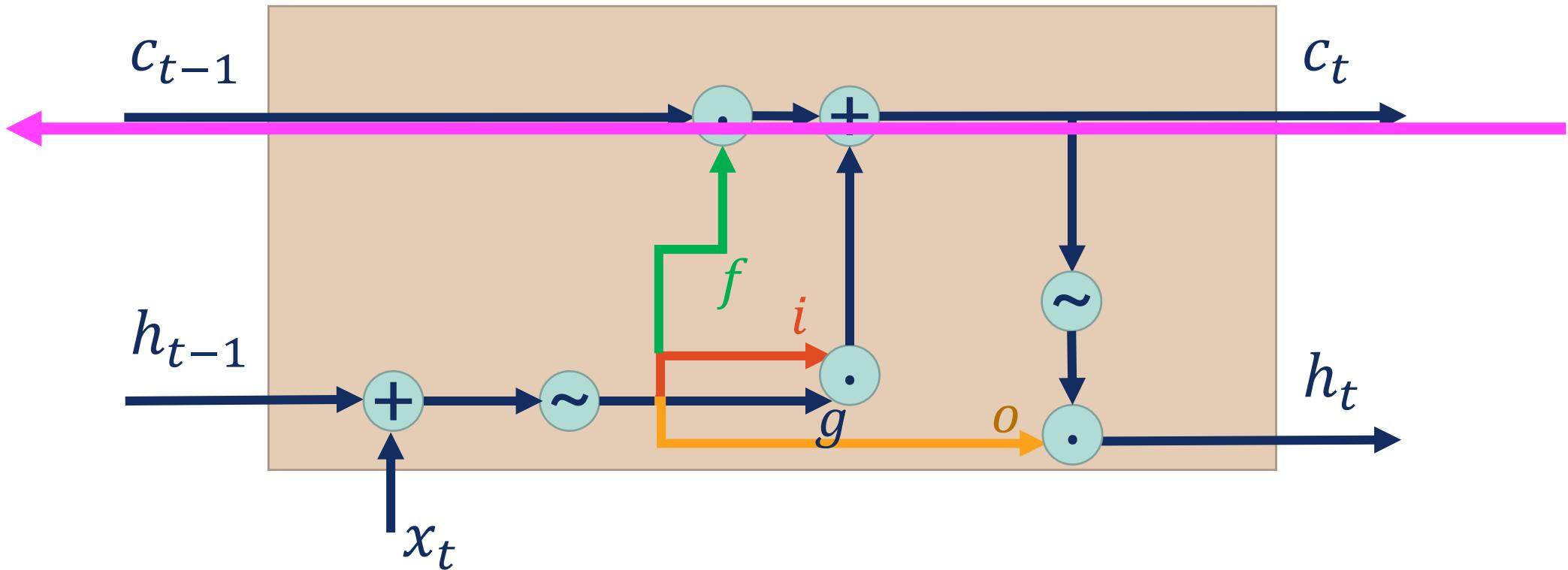


~



Вчера телефон перестал работать

Схема LSTM



$$\begin{pmatrix} g_t \\ i_t \\ o_t \\ f_t \end{pmatrix} = \begin{pmatrix} \varphi \\ \sigma \\ \sigma \\ \sigma \end{pmatrix} (W_x x_t + W_h h_{t-1} + b)$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot g_t$$
$$h_t = o_t \cdot \phi(c_t)$$

Attention

The animal didn't cross the street because it was too tired .

The diagram illustrates attention weights between two sentences. The word 'it' in the first sentence receives the most attention from all four words in the second sentence: 'animal', 'cross', 'street', and 'wide'. The word 'animal' receives the least attention from 'it'.

The animal didn't cross the street because it was too wide .

The diagram illustrates attention weights between two sentences. The word 'it' in the first sentence receives the most attention from all four words in the second sentence: 'animal', 'cross', 'street', and 'wide'. The word 'animal' receives the least attention from 'it'.

Attention

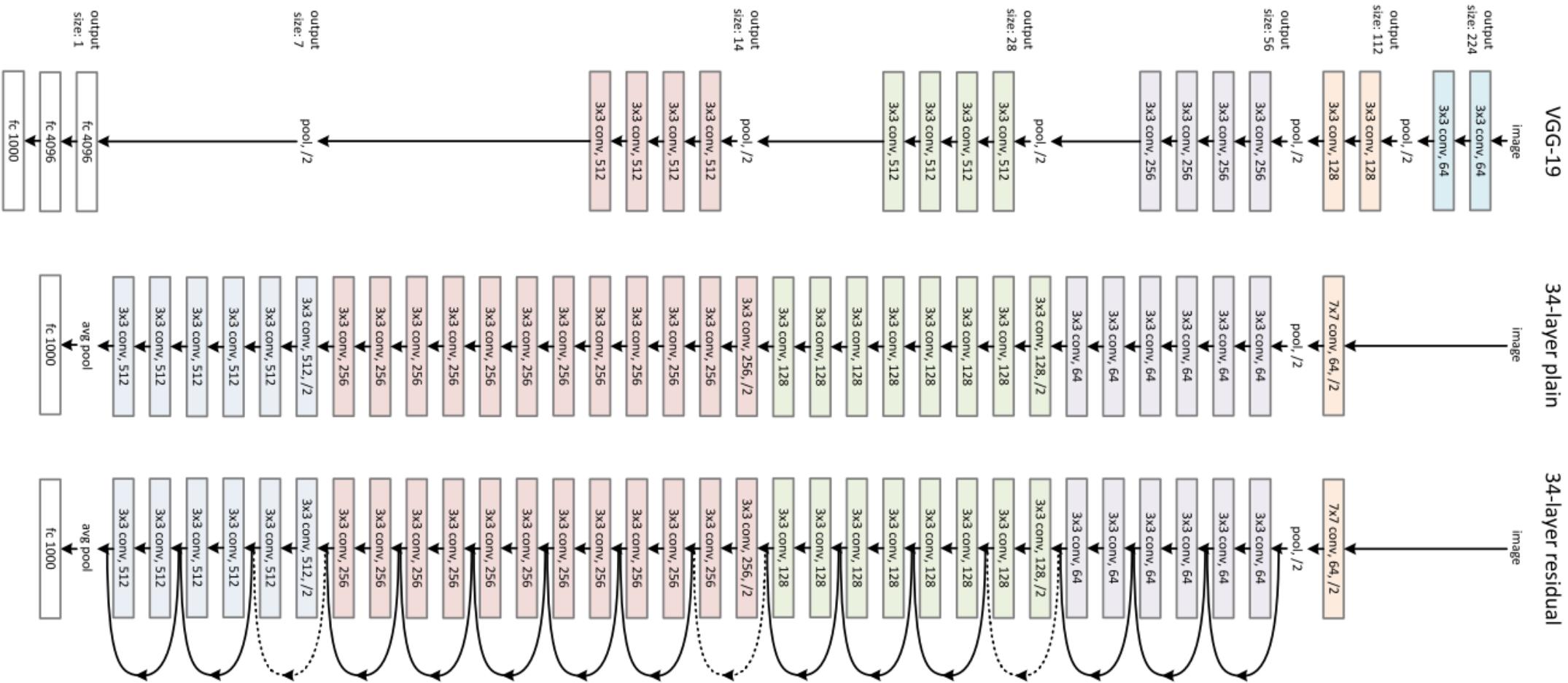
$$A_h^{(j)} = \text{softmax}(d_k^{-0.5} Q_h^{(j)} {K_h^{(j)}}^T)$$

$$M_h^{(j)} = A_h^{(j)} V_h^{(j)}$$

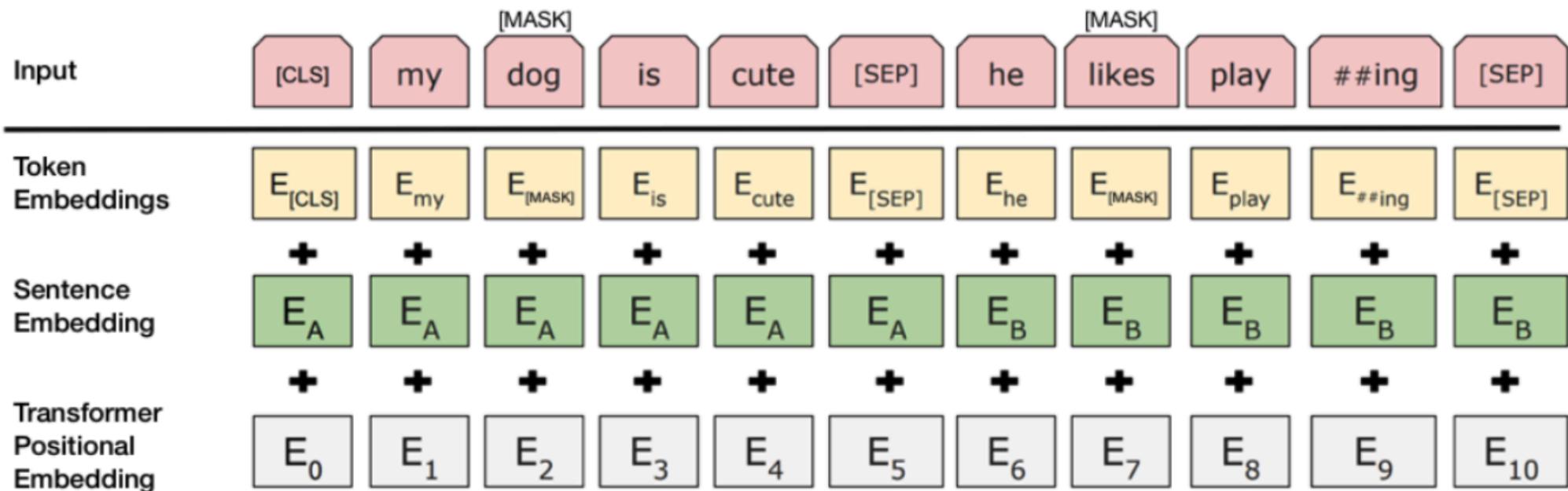
4. Сети- SOTA



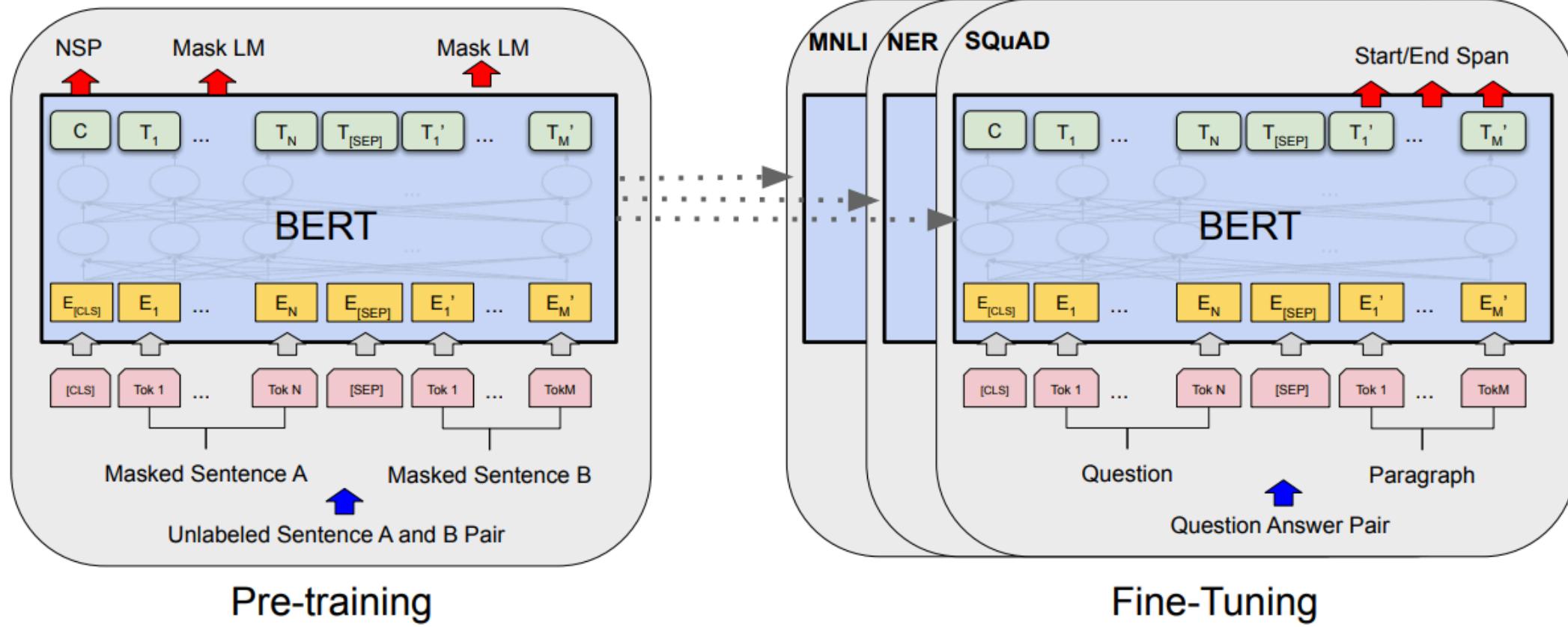
ResNet



BERT



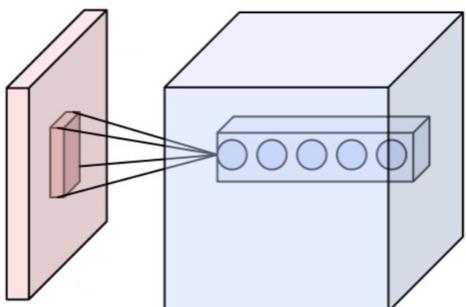
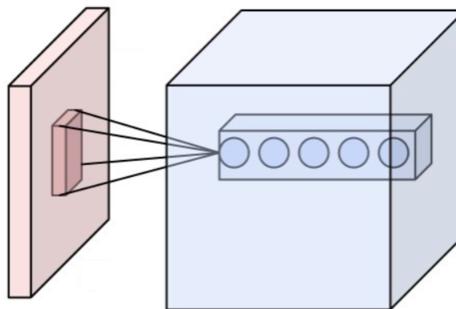
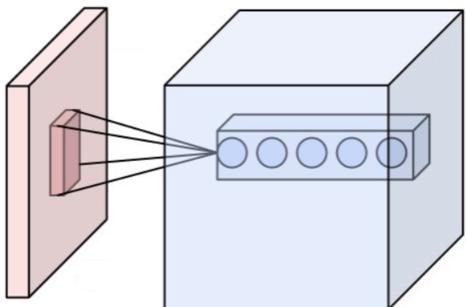
BERT



5. Практические кейсы



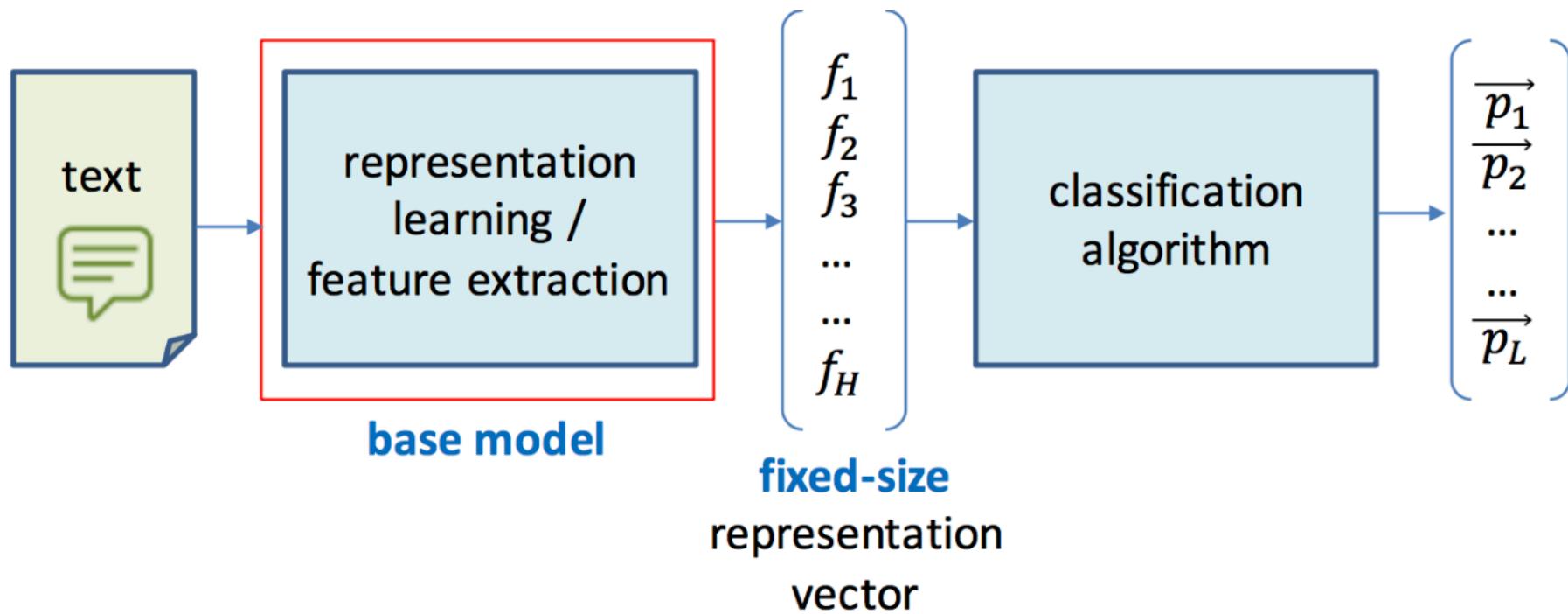
Ожидание



Как бы мне
скомбинировать 100 слоев и 10
способов регуляризации?

Реальность

Предобученные энкодеры, выдающие очень крутые признаки

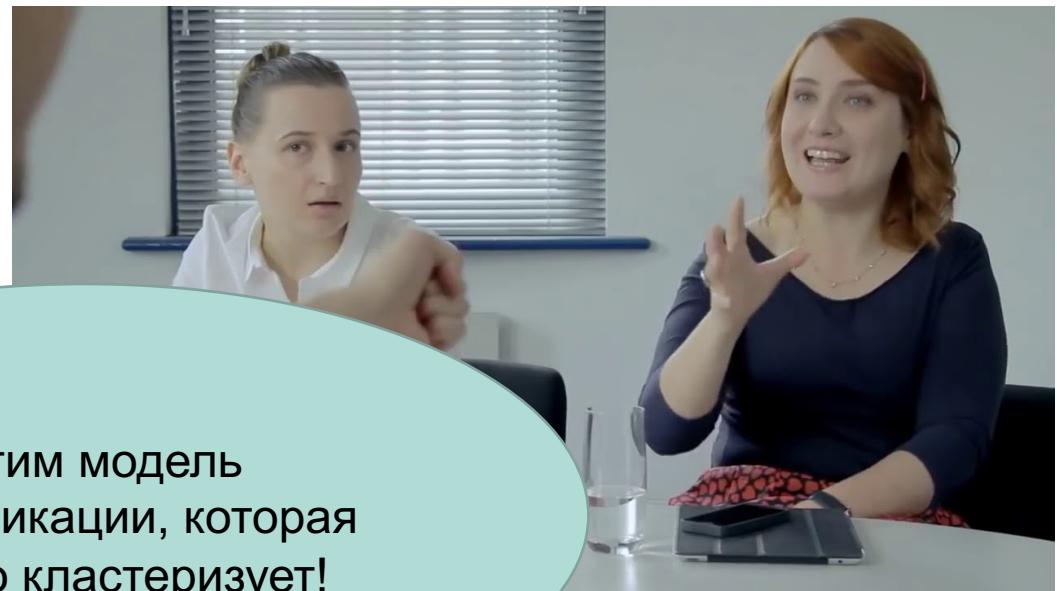


Реальность – это вызов!



Можно фантазировать на стыке
бизнеса и моделей!

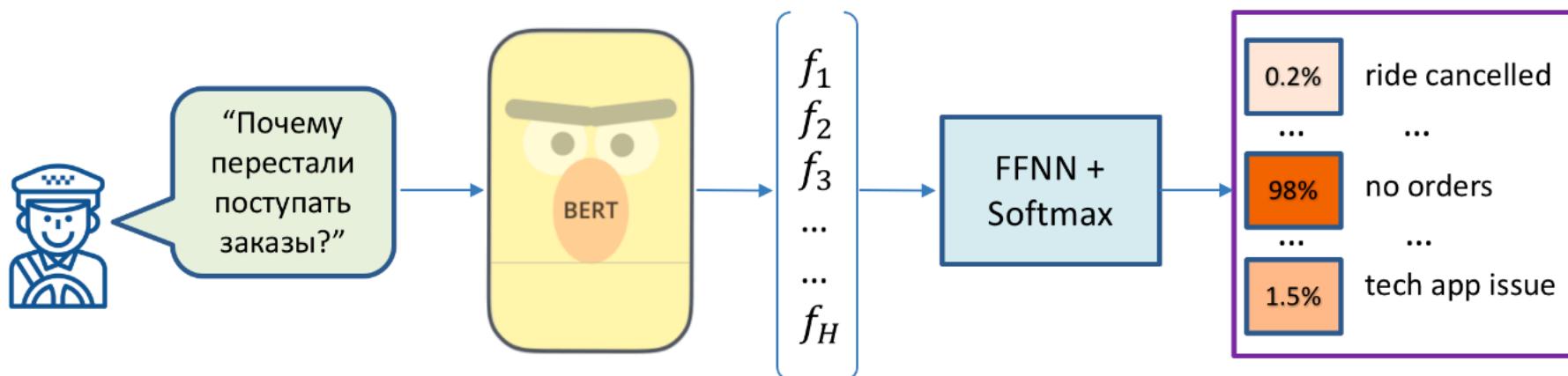
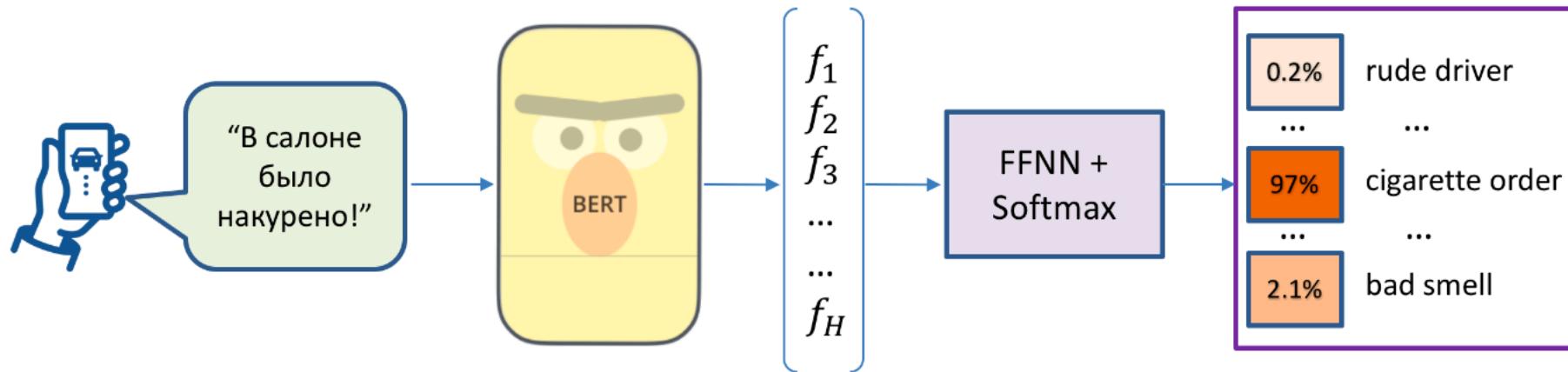
Хотим модель
классификации, которая
хорошо кластеризует!



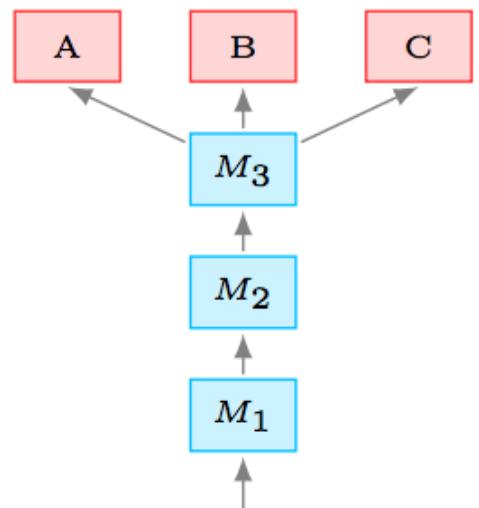
5. Практические кейсы с DL нюансами



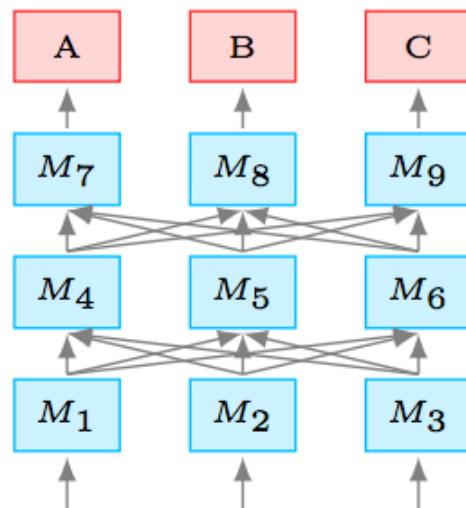
Мультитаск



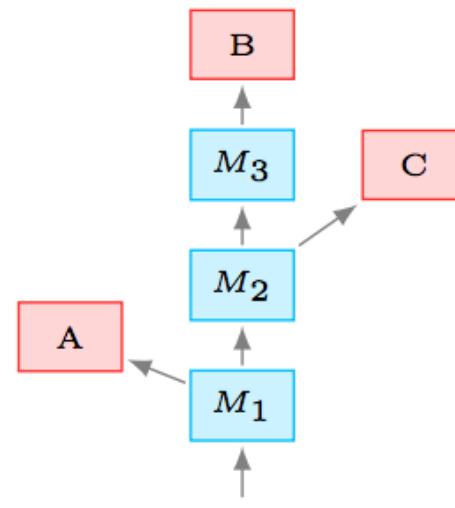
Мультитаск



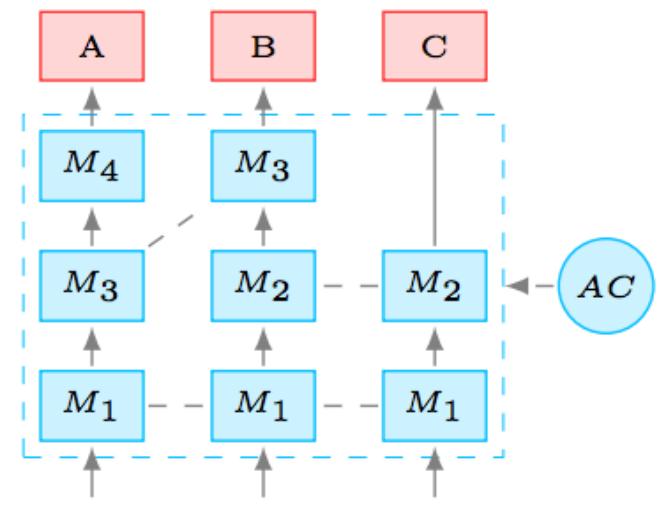
(a) Hard



(b) Soft



(c) Hierarchical



(d) This work

Иерархическая классификация

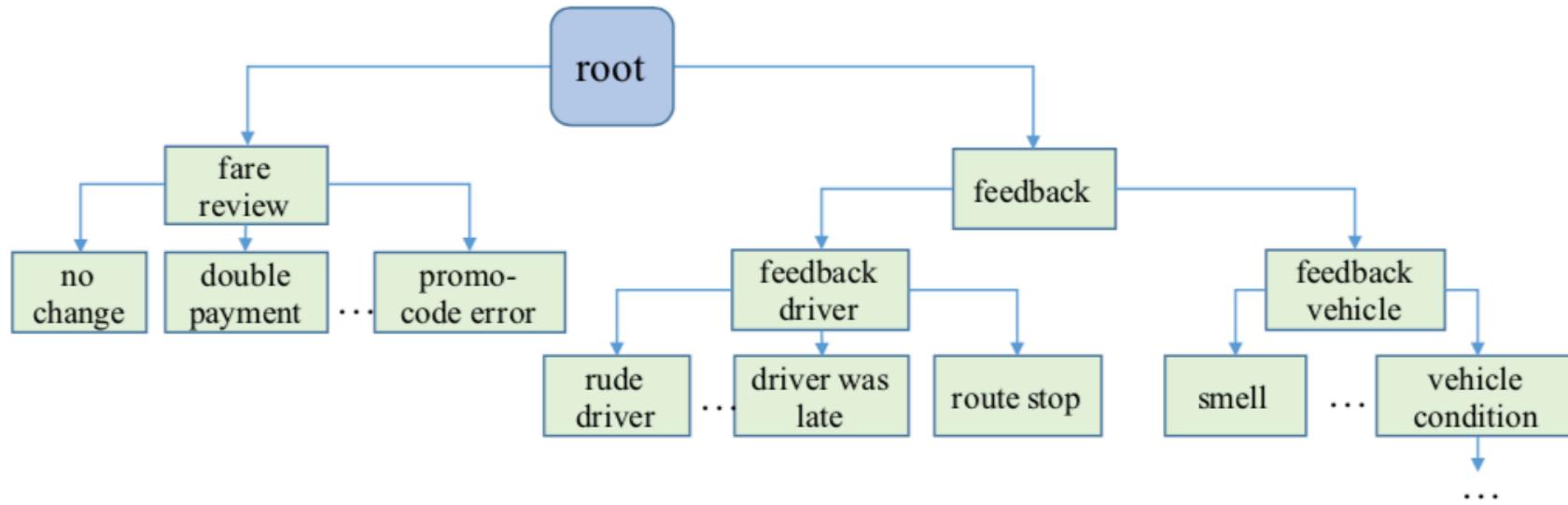


Figure 2: An example of a subtree of the tree hierarchy over classes in *Y.Taxi Riders* dataset.

Иерархическая классификация

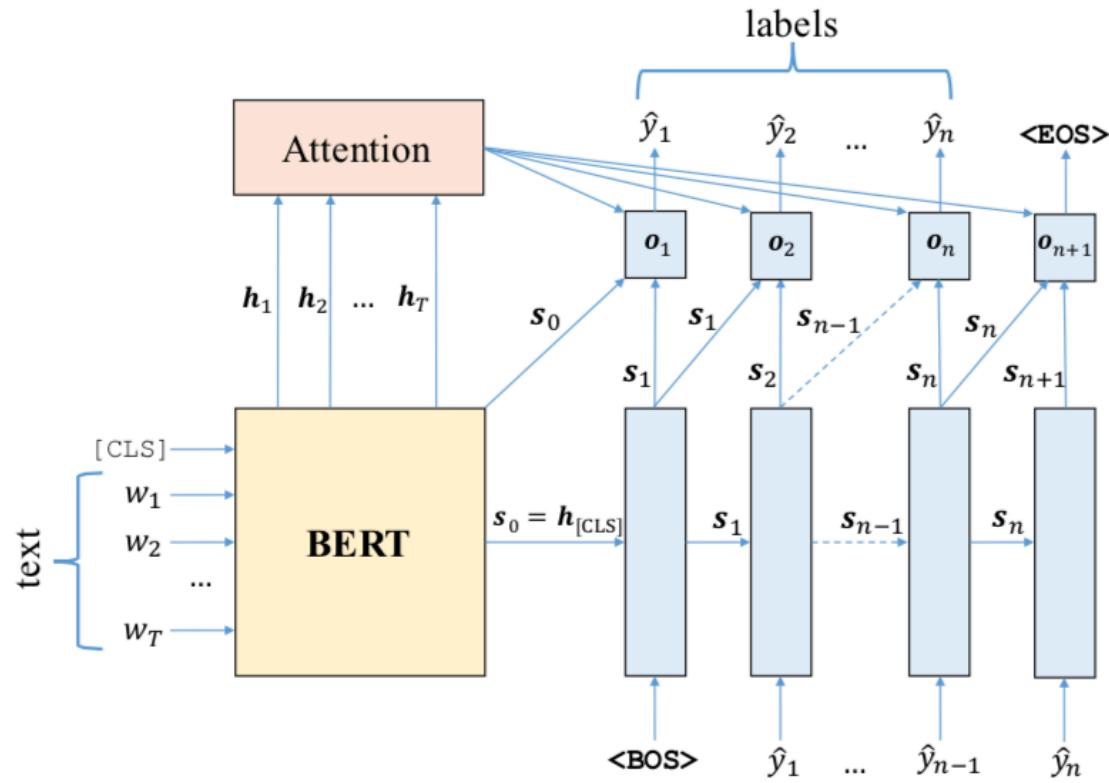
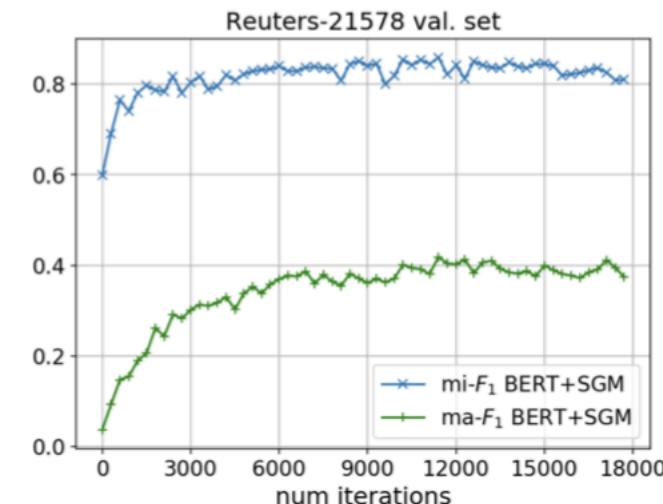
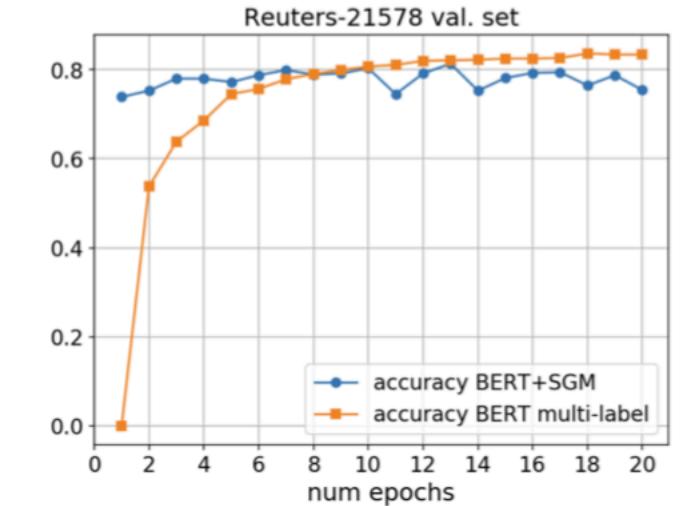
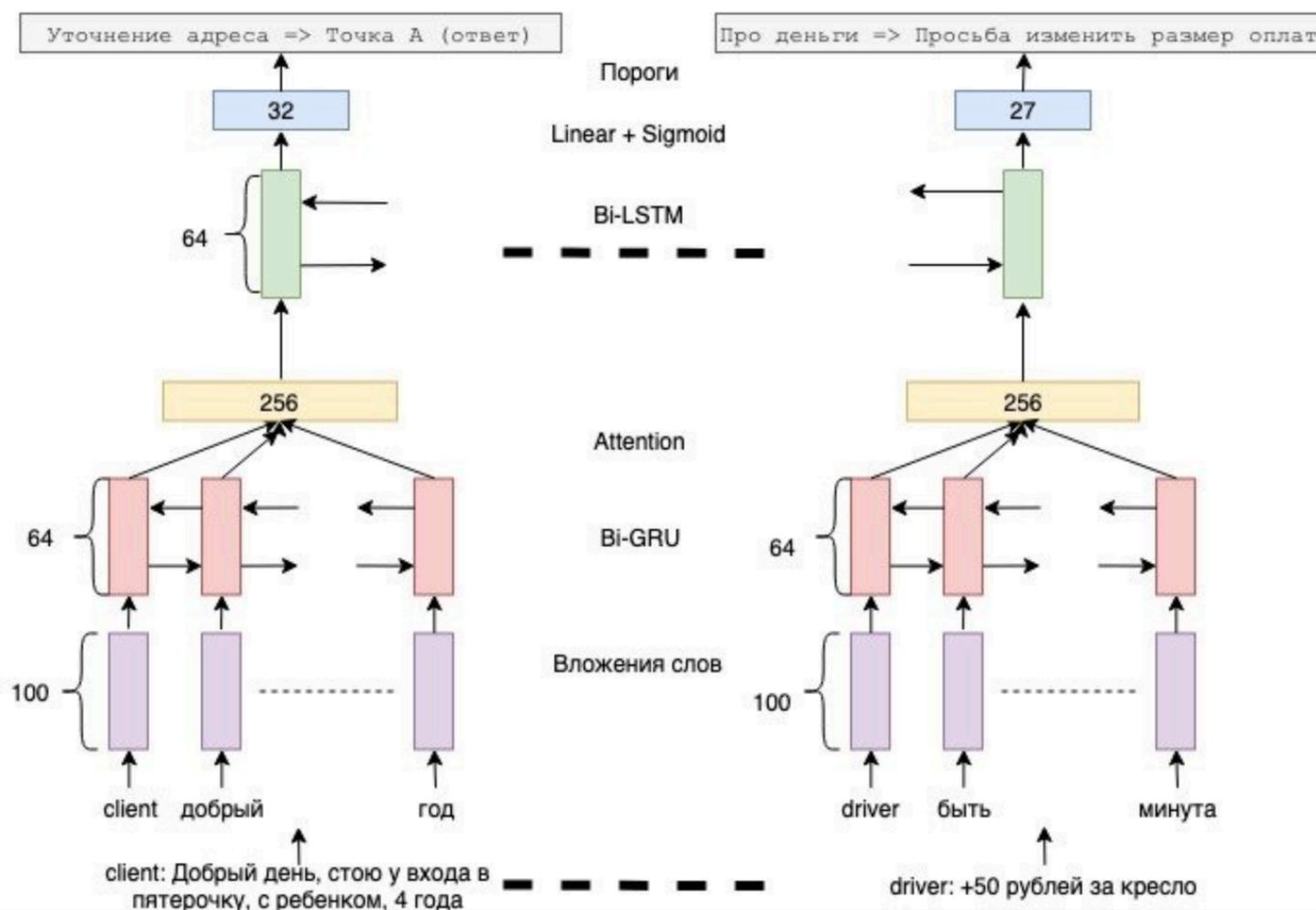


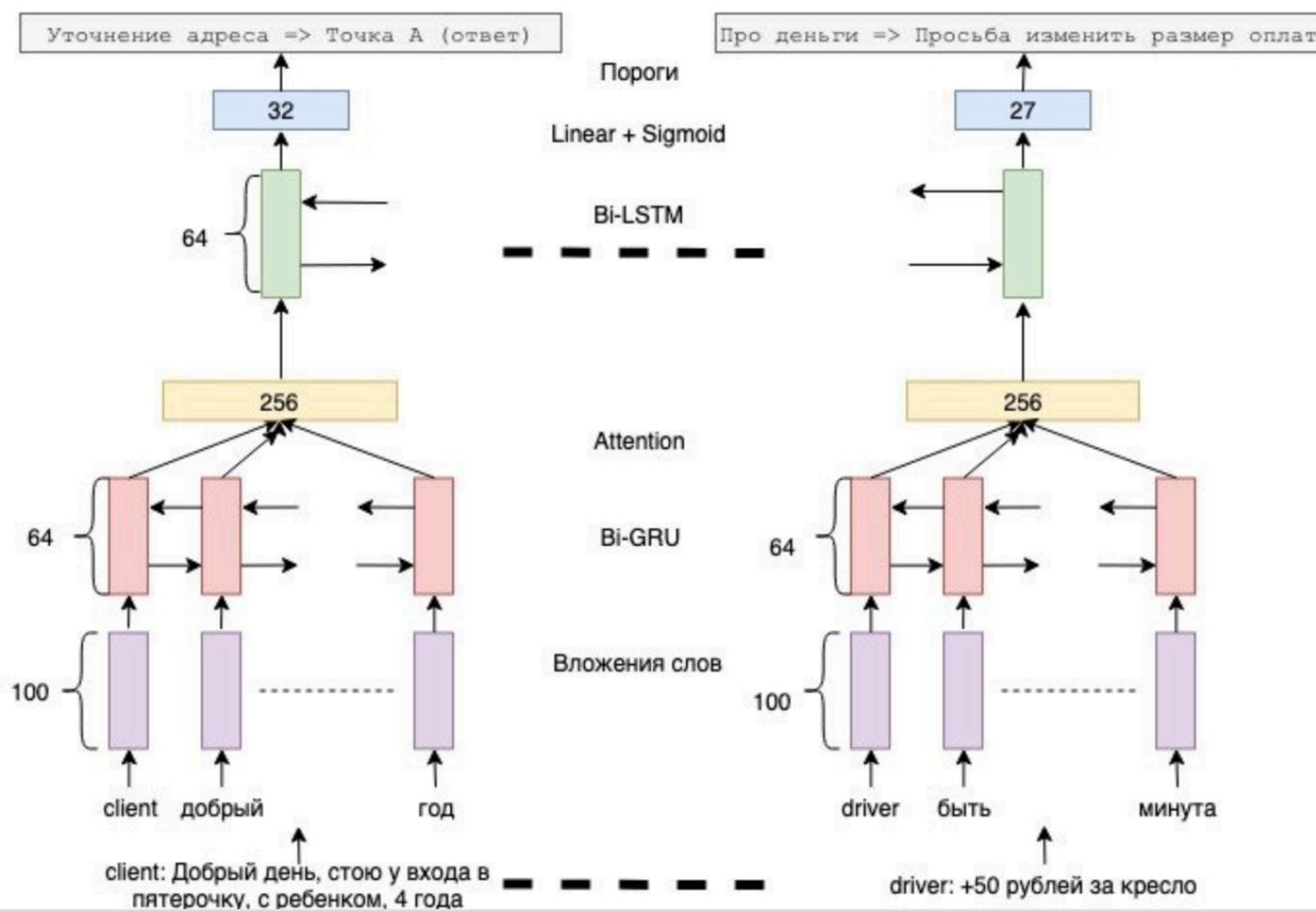
Figure 1: BERT + SGM. An overview of the model.



Диалог



Диалог



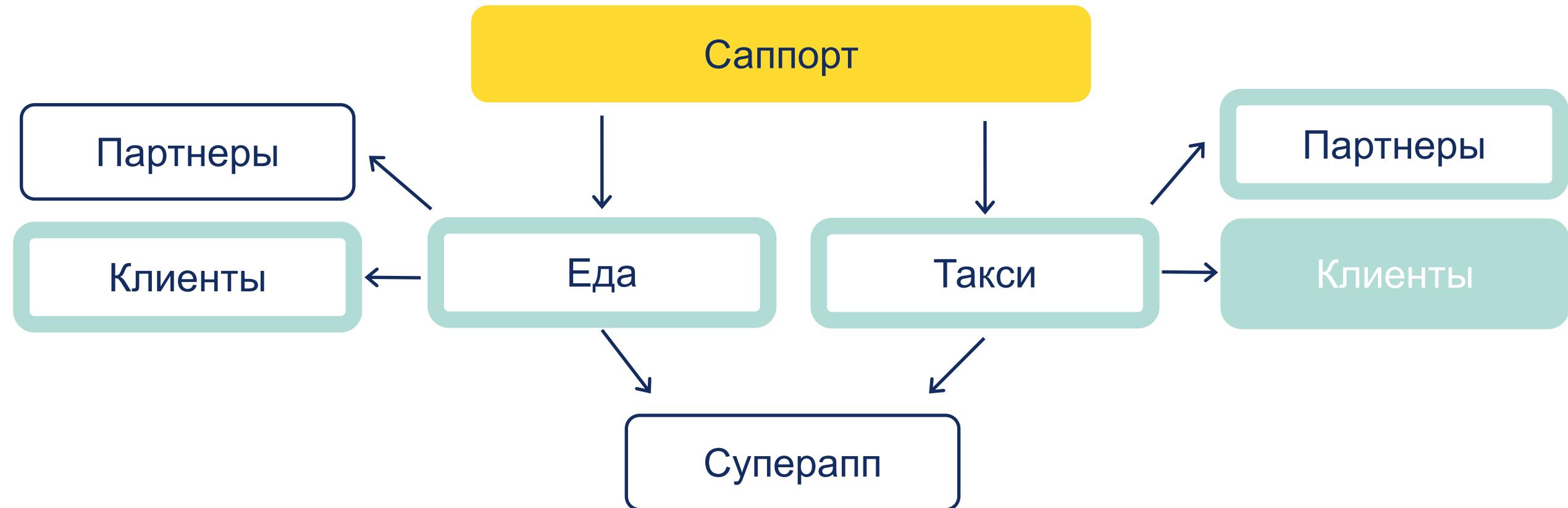
Аугментация



6. Специалист по DL – часто организует весь процесс

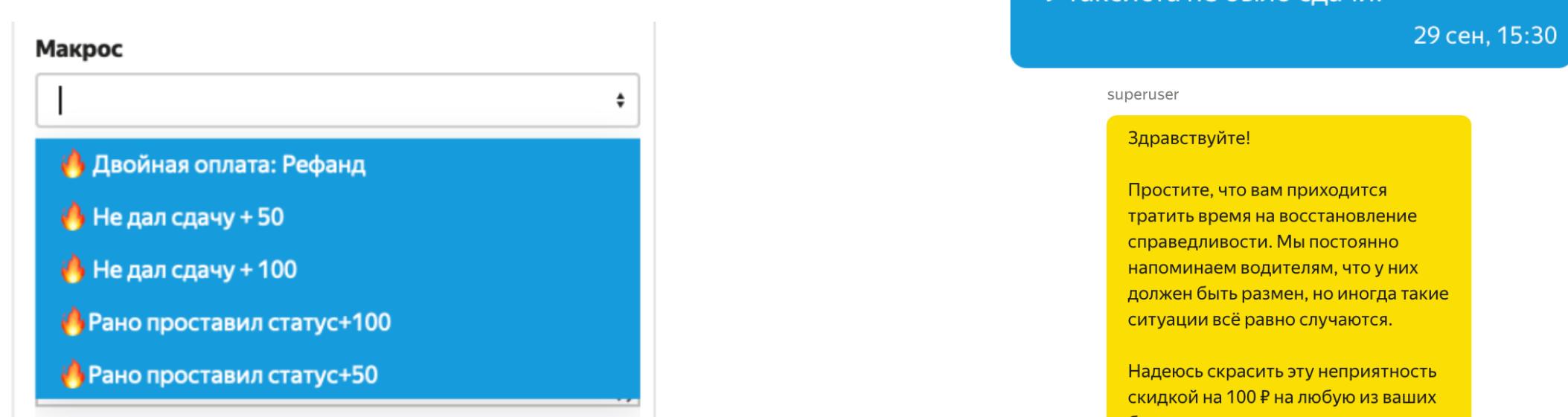


Про саппорт



Про саппорт + ML

Suggest



Autoreply

У таксиста не было сдачи!

29 сен, 15:30

superuser

Здравствуйте!

Простите, что вам приходится тратить время на восстановление справедливости. Мы постоянно напоминаем водителям, что у них должен быть размен, но иногда такие ситуации всё равно случаются.

Надеюсь скрасить эту неприятность скидкой на 100 ₽ на любую из ваших будущих поездок по промокоду bad60360351.

Спасибо, что вы с нами. Пишите, если могу ещё чем-то помочь.

29 сен, 15:38

Как работает саппорт?

Шаблон ответа

Дали промокод на скидку 100 рублей водитель взял деньги без учёта скидки! Смысл давать промокод!

27 сен, 11:43

Текст сообщения



Здравствуйте! Мне жаль, что так произошло. Видимо, водитель не заметил скидку и потому взял с вас большую сумму. Сделаем ему строгое предупреждение и возьмём под контроль отзывы о его работе. Чтобы такое не повторилось, лучше дождаться, пока водитель закроет заказ, и оплатить итоговую сумму, которую покажет приложение. Примите от нас промокод regr59175877 взамен сгоревшего. Он сделает новую поездку дешевле на 100 ₽.

Обязательно пишите, если смогу ещё чем-либо помочь.

27 сен, 11:49

Общая информация

Трек

Расчёт оффера

Маршрут

Россия, Челябинск, улица Свободы, 155Б, подъезд 4
[61.416495, 55.152495]

Россия, Челябинск, улица Академика Королёва, 52
[61.2873, 55.164186]

Страховщик

Акционерное общество «Группа Ренессанс
Страхование»

Требования

coupon

Подача

27.09.2019 в 11:22:00

27.09.2019 в 13:22:00 (+05:00, местное время)

Информация о поездке

Автореплай – первые шаги

Используем только текст для принятия решения

- › 10% автореплаев на первых линиях
- › Покрытие простых кейсов
- › Невозможность дальнейшего масштабирования

Как мне получить чек за поездку? В приложении он открывается на экране, но с ним ничего нельзя сделать. Мне нужно его напечатать.

29 сен, 16:17

superuser

Здравствуйте! Скачать чек можно в последней версии приложения. Сразу после поездки чек приходит в виде пуш-сообщения. Чтобы его скачать, нужно нажать на желтую кнопку «Поделиться» внизу.

Если вдруг пропустили уведомление, то нужно просто зайти в историю ваших поездок, выбрать там нужную, нажать на строку с суммой заказа и в окне с чеком нажать ту же кнопку «Поделиться».

Напишите, если понадобится моя помощь.

29 сен, 16:24

Но иногда этого не достаточно ...

Дали промокод на скидку 100 рублей водитель взял деньги без учёта скидки! Смысл давать промокод!

27 сен, 11:43

Как понять, что нужно давать еще промокод?

Здравствуйте! Мне жаль, что так произошло. Видимо, водитель не заметил скидку и потому взял с вас большую сумму. Сделаем ему строгое предупреждение и возьмём под контроль отзывы о его работе. Чтобы такое не повторилось, лучше дождаться, пока водитель закроет заказ, и оплатить итоговую сумму, которую покажет приложение.

Примите от нас промокод regr59175877 взамен сгоревшего. Он сделает новую поездку дешевле на 100 ₽.

Обязательно пишите, если смогу ещё чем-либо помочь.

27 сен, 11:49

Автореплай – решение



Используем информацию о поездке, прописываем бизнес-логику в коде

- › 40 тысяч автореплаев в день
- › 34% автореплаев на первых линиях
- › Статзначимо неотличающийся CSAT

Проблемы этого решения

1. Куча кода
2. ML == операционный менеджмент
3. Нет системного процесса обновления и расширения бизнес-логики

Идеальный автореплей – что хотим?

1. Неограниченно масштабироваться
2. Быстро адаптироваться
3. Улучшать систему с помощью фидбека пользователей

Автореплай – решение

Шаблон ответа

Здравствуйте! Мне жаль, что так произошло. Видимо, водитель не заметил скидку и потому взял с вас большую сумму. Сделаем ему строгое предупреждение и возьмём под контроль отзывы о его работе. Чтобы такое не повторилось, лучше дождаться, пока водитель закроет заказ, и оплатить итоговую сумму, которую покажет приложение.

Примите от нас промокод regr59175877 взамен сгоревшего. Он сделает новую поездку дешевле на 100 ₽.

Обязательно пишите, если сможете ещё чем-либо помочь.

27 сен, 11:49

Автореплай – решение

Шаблон ответа

Здравствуйте! Мне жаль, что так произошло. Видимо, водитель не заметил скидку и потому взял с вас большую сумму. Сделаем ему строгое предупреждение и возьмём под контроль отзывы о его работе. Чтобы такое не повторилось, лучше дождаться, пока водитель закроет заказ, и оплатить итоговую сумму, которую покажет приложение.

Примите от нас промокод regr59175877 взамен сгоревшего. Он сделает новую поездку дешевле на 100 ₽.

Обязательно пишите, если сможете ещё чем-либо помочь.

27 сен, 11:49

Тематика

Бизнес-логика

Автореплай – решение

Шаблон ответа

Здравствуйте! Мне жаль, что так произошло. Видимо, водитель не заметил скидку и потому взял с вас большую сумму. Сделаем ему строгое предупреждение и возьмём под контроль отзывы о его работе. Чтобы такое не повторилось, лучше дождаться, пока водитель закроет заказ, и оплатить итоговую сумму, которую покажет приложение.

Примите от нас промокод regr59175877 взамен сгоревшего. Он сделает новую поездку дешевле на 100 ₽.

Обязательно пишите, если смогу ещё чем-либо помочь.

27 сен, 11:49

Тематика

Пересмотр стоимости поездки

Промокод не применился

Бизнес-логика

- `complete_rides >1`
- `success_order_flg is true`
- `paid_cancel_tag is false`
- `final_transaction_status != 'hold_fail'`
- `coupon_used is true`
- `coupon_use_value = 100`

ML

Поддержка

Бывают забавные кейсы ...

Очень понравилось всё!
Машина приехала
вовремя, хороший
водитель, приемлемые
цены! Спасибо

Артём

Здравствуйте! Спасибо,
что рассказали. С
водителем свяжемся, а
также передадим
информацию в таксопарк.
Если такое повторится, он
больше не сможет
выполнять наши заказы.



7. Выводы



Что нужно зафиксировать в конце лекции?

- Минимум теории про сети, перечень нужных понятий
- Какие задачи в индустрии решают с помощью DL
- Что нужно знать, чтобы успешно их решать
- Где возникают нюансы в применении методов
- Зачем и где саморазвиваться?