

## Возможные причины переобучения:

- 1 линейная зависимость (мультиколлинеарность) признаков:  
пусть построен классификатор:  $a(x, w) = \text{sign}\langle w, x \rangle$ ;  
мультиколлинеарность:  $\exists u \in \mathbb{R}^{n+1}: \forall x \quad \langle u, x \rangle \equiv 0$ ;  
тогда  $\forall \gamma \in \mathbb{R} \quad a(x, w) = \text{sign}\langle w + \gamma u, x \rangle$
- 2 слишком мало объектов; слишком много признаков;

## Проявления переобучения:

- 1 слишком большие веса  $|w_j|$  разных знаков;
- 2 неустойчивость классификаций  $a(x, w)$  относительно погрешностей измерения признаков;
- 3  $Q(X^\ell) \ll Q(X^k)$ ;

Штраф за увеличение нормы вектора весов:

$$\widetilde{\mathcal{L}}_i(w) = \mathcal{L}_i(w) + \frac{\tau}{2} \|w\|^2 = \mathcal{L}_i(w) + \frac{\tau}{2} \sum_{j=1}^n w_j^2 \rightarrow \min_w.$$

Градиент:

$$\nabla \widetilde{\mathcal{L}}_i(w) = \nabla \mathcal{L}_i(w) + \tau w.$$

Модификация градиентного шага:

$$w := w(1 - h\tau) - h\nabla \mathcal{L}_i(w).$$

Подбор параметра  $\tau$  — по скользящему контролю

- В методе стохастического градиента необходимы различные эвристики для улучшения сходимости и получения лучшего решения.
- *Регуляризация* решает проблему мультиколлинеарности и снижает риск переобучения.