# Cvičení 8b

## Výběr modelu

Jan Přikryl

ČVUT FD

14. dubna 2020

# Příklad 1

## Hřebenová regrese

Použijeme sadu statistik z baseballu. První sloupec v `islr_hitters.csv` jsou jména hráčů, proto `'ReadRowNames'`. U platů občas není uveden plat, je tam `'NA'`. Tyto hodnoty je třeba převést na NaN, proto `'TreatAsEmpty'`

```
hitters = readtable('islr_hitters.csv', 'ReadRowNames', true,
                    'TreatAsEmpty', 'NA');
```

Informace o datech

```
summary(hitters)
size(hitters)
```

# Příklad 1

## Pokračování

Zjistíme, kolik tam máme hráčů bez uvedeného platu a odstraníme všechny neúplné záznamy

```
sum(isnan(hitters.Salary))
missing = ismissing(hitters);
hitters2 = hitters(~any(missing,2),:);
size(hitters2)
sum(isnan(hitters2.Salary))
```

Bohužel, Matlab neumí zpracovávat hřebenovou regresí data, uložená v tabulce, vstupem funkce `ridge()` je matice regresorů a vektor odezev. Provedeme proto úpravu formátu použitých dat.

# Příklad 1

## Pokračování

Vše číselné, tedy `League`, `NewLeague` a `Division` musíme konvertovat na čísla:

```
hitters2.League = double(categorical(hitters2.League))-1;
hitters2.Division = double(categorical(hitters2.Division))-1;
hitters2.NewLeague = double(categorical(hitters2.NewLeague))-1;
```

Hřebenová regrese potřebuje vektor výstupu a matici vstupu, ve vstupech nesmí být plat:

```
X = table2array(hitters2);
X(:,end-1) = []; % Smazat platy
y = hitters2.Salary;
```