Ridge Regression

Genji Ohara

June 9, 2023

1 What does the code mean?

1.1 Preamble

```
module Ridge where

import Numeric.LinearAlgebra
import Prelude hiding ((<>))

type Vec = Vector R
type Mat = Matrix R
```

1.2 Adding Bias

$$\tilde{\boldsymbol{X}} \triangleq \begin{bmatrix} 1 & \\ 1 & \\ \vdots & \boldsymbol{X} \\ 1 & \end{bmatrix}, \quad \tilde{\boldsymbol{w}} \triangleq \begin{bmatrix} w_0 \\ \boldsymbol{w} \end{bmatrix}$$

```
addBias :: Mat → Mat
addBias x = fromColumns $ [bias] ++ (toColumns x)
    where bias = vector $ take (rows x) [1,1..]
```

1.3 Prediction

$$\hat{\boldsymbol{y}} = \begin{bmatrix} w_0 \\ w_0 \\ \vdots \\ w_0 \end{bmatrix} + \boldsymbol{X}\boldsymbol{w} = \tilde{\boldsymbol{X}}\tilde{\boldsymbol{w}}$$

```
predict :: Vec → Vec → R
predict w x = w <.> (vector $ [1.0] ++ toList x)
```

1.4 Fitting

We minimize the objective:

$$E(\tilde{\boldsymbol{w}}) = \|\boldsymbol{y} - \tilde{\boldsymbol{X}}\tilde{\boldsymbol{w}}\|^2 + \lambda\|\tilde{\boldsymbol{w}}\|^2$$

Gradient:

$$\nabla E(\tilde{\boldsymbol{w}}) = 2\left[\left(\tilde{\boldsymbol{X}}^T\tilde{\boldsymbol{X}} + \lambda I\right)\tilde{\boldsymbol{w}} - \tilde{\boldsymbol{X}}^T\boldsymbol{y}\right]$$

Therefore

$$\underset{\tilde{\boldsymbol{w}}}{\operatorname{argmin}}\ E(\tilde{\boldsymbol{w}}) = \left(\tilde{\boldsymbol{X}}^T\tilde{\boldsymbol{X}} + \lambda I\right)^{-1}\tilde{\boldsymbol{X}}^T\boldsymbol{y}$$

```
fit :: Mat → Vec → R → Vec
fit x_til y lambda = (inv a) #> ((tr x_til) #> y)
  where a = (tr x_til) <> x_til + (scale lambda $ ident $ cols x_til
      )
```