product rule of probability: p(h|y)p(y) = p(h)p(y|h) = p(h, y)

A screenshot of a math test

Description automatically generated

A screenshot of a math test

Description automatically generatedA screenshot of a math test

Description automatically generated

Var(x) = E[(X-E[X])^2]=E[X^2]-E[X]^2

A math equations and formulas

Description automatically generated with medium confidence

GaussianA black and white math equation

Description automatically generated with medium confidence

A math equations and formulas

Description automatically generated with medium confidenceA math equations and formulas

Description automatically generated with medium confidence

A screenshot of a math problem

Description automatically generated

Linear regression

A black and white image of a triangle and a triangle with a triangle and a triangle with a triangle and a triangle with a triangle and a triangle with a triangle and a triangle with a triangle and

Description automatically generated

Set derivate = 0 solve for w

A white board with red text and numbers

Description automatically generatedA math equations and formulas

Description automatically generated with medium confidence

Offset – b

A math equations and symbols

Description automatically generated with medium confidence

LOO Cross validation

Unbiased estimate, picking degree

k-fold: much biased (much less data used): n(k-1)/k

B/V trade-off

A math equations and formulas

Description automatically generated with medium confidence

A math equations with green and purple text

Description automatically generated with medium confidence

Ridge Regression

When x\_i in R^d and d>n, the objective function is flat in some directions

A black and white math equation

Description automatically generated with medium confidence

A screenshot of a math equation

Description automatically generated

A math equations on a white background

Description automatically generated

Lambda increase, train error decrease. While test error has U shape

Lasso regression

1 model selection, choose lambda based on cv error.

2 feature selection

3 retrain the sparse model and let lambda = 0.

**Gradient descent**

For linear regression:

A math equations with orange and black text

Description automatically generated

Eta too big, does not converge.

Eta too small converges slow.

For ridge regression:

A math equations and formulas

Description automatically generated with medium confidenceA math equations on a white background

Description automatically generated

A group of math equations

Description automatically generated

Sub gradient: a function is non-smooth if it is not differentiable everywhere.

A math equations on a white background

Description automatically generated

32: cdba;bbbcb

1:cdb(0.80)b; 6:a(train)(4)(more training data, reduce model complexity)(a*c*d)

11:dbcdb; (n\*m, )b(15%)ab;

21: 2?c