UNIVERSITY OF OSLO

Faculty of mathematics and natural sciences

Exam in: STK-IN4300/STK-IN9300 — Statistical learning methods in Data Science

Day of examination: Monday, December 17th, 2018

Examination hours: 14.30 – 18.30

This problem set consists of 4 pages.

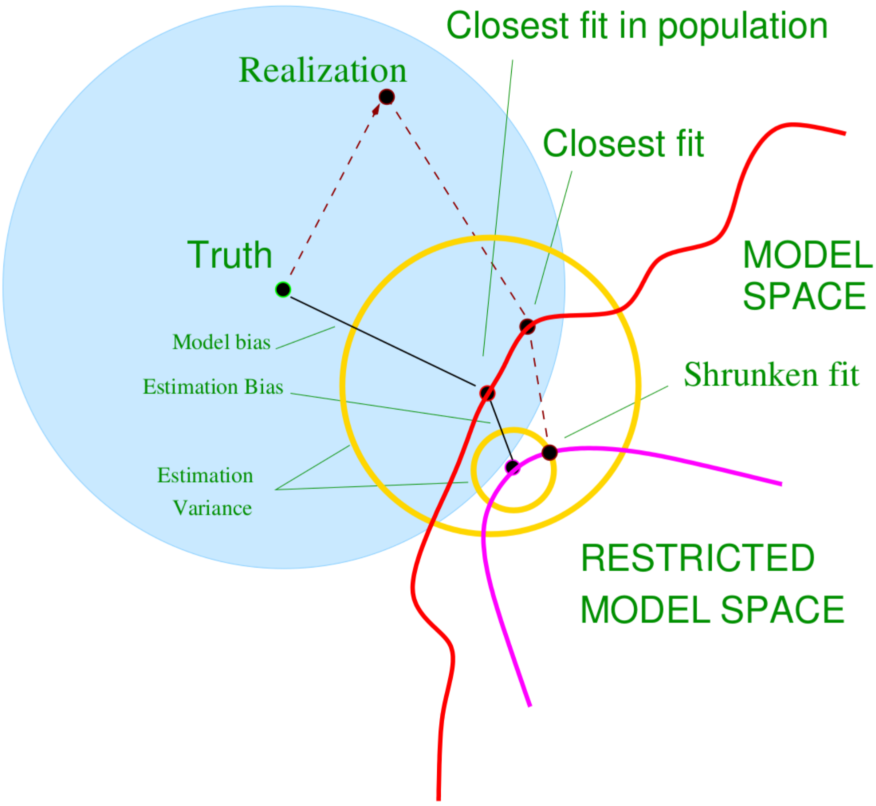
Appendices: None.

Permitted aids: None.

Please make sure that your copy of the problem set is complete before you attempt to answer anything.

Problem 1 Penalized regression

Consider the following figure from the text book (Hastie, Tibshirani & Friedman, 2009, The Elements of Statistical Learning):



a

Using the figure above, explain the concept of bias-variance trade-off.

(Continued on page 2.)

b

Show analytically the same concept of the point above by mathematically comparing bias and variance of the ordinary least square estimator and of the ridge estimator.

Problem 2 Bootstrapping for model evaluation

Consider the following procedure to estimate the prediction error:

1. generate B bootstrap samples $z_1, \dots, z_B$, where $z_b = \{(y_1^*, x_1^*), \dots, (y_N^*, x_N^*)\}$, $b = 1, \dots, B$ and (y_i^*, x_i^*), $i = 1, \dots, N$, is an observation sampled from the original dataset;

2. apply the prediction rule to each bootstrap sample to derive the predictions $\hat{f}_b^*(x_i)$, $b = 1, \dots, B$;

3. compute the error for each point, and take the average,

$$\widehat{\mathrm{Err}}_{\mathrm{boot}} = \frac{1}{B}\sum_{b=1}^{B}\frac{1}{N}\sum_{i=1}^{N} L(y_i, \hat{f}_b^*(x_i)).$$

a

Explain why this procedure is incorrect and suggest a different way to proceed which still uses a bootstrap approach.

b

Describe the 0.632 `bootstrap` and the 0.632+ `bootstrap` procedures, explaining in particular the rationale behind their construction.

Problem 3 Smoothing splines

Consider the following problem: among all functions $f(x)$ with two continuous derivatives, find one that minimizes the penalized residual sum of squares

$$RSS(f, \lambda) = \sum_{i=1}^{N}(y_i - f(x_i))^2 + \lambda \int \{f''(t)\}^2 dt \qquad (1)$$

where $\lambda \geq 0$.

a

Define the role of the penalization term $\lambda \int \{f''(t)\}^2 dt$ in relation to its specific form, and discuss what happens when the smoothing parameter λ varies.

(Continued on page 3.)

b

The solution of minimization problem (1) the can be written as a natural spline

$$f(x) = \sum_{j=1}^{N} N_j(x)\theta_j.$$

Rewrite (1) as a function of θ (i.e., $RSS(\theta, \lambda)$) and use its solution to show that a smoothing spline for a fixed λ is a linear smoother (linear operator). Use it to define the effective degrees of freedom of a smoothing spline.

Problem 4 Bagging

a

Describe bagging, mentioning at least one advantage with respect to a single tree and a disadvantage with respect to a boosted tree model.

b

Consider a classification problem and how to aggregate the results of the single trees in a bagging classifier. The aggregation can be done by looking at the estimated classes or at the class-probability estimates. Show with a simple example that the two procedures can produce different results in terms of classification of an observation.

Problem 5 Boosting

a

Show that the additive expansion produced by AdaBoost is estimating one-half the log-odds of $P(Y = 1|X = x)$, where Y is the binary response and X the input matrix.

b

Consider the following algorithm,

---

1. initialize the estimate, e.g., $f_0(x) = 0$;
2. for $m = 1, \dots, m_{\text{stop}}$,
   - 2.1 compute the negative gradient vector, $u_m =$ **OMITTED**;
   - 2.2 fit the base learner to the negative gradient vector, $h_m(u_m, x)$;
   - 2.3 update the estimate, $f_m(x) = f_{m-1}(x) + \nu h_m(u_m, x)$.

(Continued on page 4.)

3. final estimate, $\hat{f}_{m_{\text{stop}}}(x) = \sum_{m=1}^{m_{\text{stop}}} \nu h_m(u_m, x)$

Name the specific boosting algorithm and write the complete formula in point 2.1 (where the word **OMITTED** is) for a generic loss function $L(y, f(x))$.

THE END