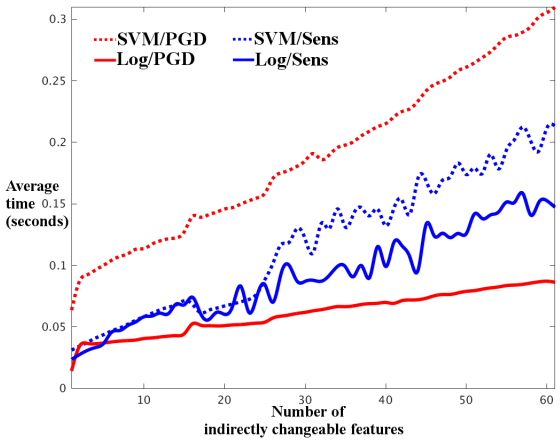
Supplement of "A budget-constrained inverse classification framework for smooth classifiers"

Michael T. Lash* Qihang Lin† Nick Street† Jennifer G. Robinson‡

1 Supplemental Experiment: Time Complexity of H

Figure SF 1 shows the average time complexity of using H, the indirectly changeable feature estimator, in the inverse classification process as the size of I increases. The experiment is motivated by our analysis of the time complexity of H in Section III.A.1.



SF 1: The average time complexity of H as the indirectly changeable set I increases on the ARIC dataset.

2 Supplementary Tables

These tables show the unchangeable, indirectly changeable, and directly changeable features for each of our two freely available datasets. The indirectly changeable features table includes the *sigma* parameter used with the kernel regression indirect feature estimator. Note, however, that we have not included the β coefficients and intercept values for the LASSO indirect feature estimator. This is because there are a rather large number of these. If the reader finds that they would like to have these, please contact the corresponding author and they will be provided.

*Department of Computer Science, University of Iowa
†Department of Management Science, University of Iowa
‡Department of Epidemiology, University of Iowa

Feature Name
School Attended, Sex, Age, Address, Size of family, Parent's cohabitation status, Mother's education, Father's education, Mother's job= "At Home", Mother's job="Health", Mother's job="Other", Mother's job="Services", Mother's job="Teacher", Father's job="Teacher", Father's job="Other", Father's job="Services", Father's job="Health", Father's job="At Home", Reason for school="Course", Reason for school="Other", Reason for school="home", Reason for school="Reputation", Guardian="Mother", Guardian="Father", Guardian="Other", Time spent traveling to school

Table 1: Unchangeable features for the Student Performance dataset.

Feature Name: σ
Extra-curricular activities: 1.5, Higher education aspirations: 1.0, In a romantic relationship: 1.5, Free time after school: 1.0

Table 2: Indirectly changeable features and learned kernel regression σ parameters for the Student Performance dataset.

$\mathbf{c^+/c^-}$	**Feature:Cost**
c^+	Study time: 7, Paid tutoring: 8
c^-	Time out with friends: 6, Weekday alcohol: 3, Weekend alcohol: 6, Absences from class: 5

Table 3: Directly changeable variables for the Student Performance dataset.

Feature Name
Insulin (uu-ml), Height (cm), Age, Peripheral Artery Disease, Peripheral Artery Disease (definition 2), Plaque/shadowing in either internal, Plaque in either internal carotid, Cholesterol lowering med (last 2 weeks), Hypertension (definition 5), Education level, Diabetes, Age when menopause began, Menopause status, Ever smoked cigarettes, High blood pressure med (past 2 weeks), Agina-chest pain med (past 2 weeks), Heart rhythm control med (past 2 weeks), Heart failure med (past 2 weeks), Blood thinning med (past 2 weeks), Blood sugar med (past 2 weeks), Stroke med (past 2 weeks), Walking leg pain med (past 2 weeks), Headache or cold med (past 2 weeks), Pain meds (past 2 weeks), Gender, Race, Years smoked cigarettes

Table 4: Unchangeable features for the ARIC CVD dataset.

Feature Name: σ
BMI (Body Mass Index): .5, Recalibrated HDL cholesterol (mg/dl): .5, Re-calibrated LDL cholesterol (mg/dl): .5,Total cholesterol (mmol/L): .5, Total triglycerides (mmol/L): .5, 2nd and 3rd systolic blood pressure (avg.): .5, 2nd and 3rd systolic blood pressure (avg.) Num 2: .5, Waist girth (cm): .5, Hip girth (cm): .5, Heart rate: .5, White blood count: .5

Table 5: Indirectly changeable features and learned kernel regression σ parameters for the ARIC CVD dataset.

Feature Name: σ
Apolipoprotein AI(mg-dl): .5, Apolipoprotein B (mg-dl): .5, Apolp(A) Data (ug-ml): .5, Ankle-brachial index (Def 4): .5, FV(1)/FVC Predicted (%): .25, FEV(1) (L): .5, FVC (L): .5, Hematocrit: .5, Hemaglobin: .5, Platelet count: .5, Neutrophils: .5, Neutrophil bands: .5, Lymphocytes: .5, Monocytes: .5, Eosinophils: .5, Basophils: .5, APTT Value: .5, VIII: C Value: .5, Fibrinogen Value: .5, VII Value: .5, ATIII Value: .5, Protein: C Value: .5, VWF Value: .5, Cornell voltage (uV): .5, Waist-hip ratio: .5, Vegetable fat (% kcal): .5, Carbs (% kcal): .5, Alcohol (% kcal): .5, Omega fatty acid (g): .5, Calf girth (cm): .5, Subcaps measure 2 (mm): .5, Triceps measure 2 (mm): .5, Uric acid (mg-dl): .5, Total protein (gm-dl): .5, Albium (gm-dl): .5, Phosphorus (mg-dl): .5, Magnesium (meq-l): .5, Calcium (mg-dl): .5, Urea nitgrogen (mg-dl): .5, Potassium (mmol-l): .5, Sodium (mmol-l): .5, Creatinine (mg-dl): .5, Weight (lb): .5,Total fat (% kcal): .5, Saturate fatty acid (% kcal): .5, Protein (% kcal): .5, Polyunsaturated fatty acid (% kcal): .5, Monounsaturated fatty acid (% kcal): .5, Total fat (g): .25

Table 6: Indirectly changeable features and learned kernel regression σ parameters for the ARIC CVD dataset.

$\mathbf{c}^+/\mathbf{c}^-$	Feature:Cost
c^+	Dark or grain breads: 3, Peanut butter: 4, Nuts: 5, Other(prunes,avocado): 5, Vegetables: 6, Fruit: 6, Fiber: 7, Vegetable fat: 5, Polyunsaturated fat: 5
c^-	Liver: 8, White carbs: 6, Fish: 9, Cereal: 4, Cigarettes: 9, Caffeine: 7, Carbs: 7, Cholesterol: 6, Sodium: 7, Animal fat: 7, Saturated fat: 6
c^+/c^-	Exercise hours: 10, Alcohol: 9

Table 7: Directly changeable variables for the ARIC CVD dataset.