Slice Models in GAMS

Michael C. Ferris

Meta M. Voelker

Outline

- Define slice models
- Example: Breast Cancer Diagnosis
  - Cross-Validation
  - Linear Classification (LP)
  - Support Vector Machine (QP)
- Introduce slice interface (GAMS/DEA)
- Train and test models
- Show results on WDBC database

Slice Models

Slice model = group of mathematical programs with the same model but different data

$$
\begin{array}{lll}
\min_x & f^k(x) & \rightarrow \text{objective slice} \\
\text{subject to} & A^k x = b^k & \rightarrow \text{slice constraints} \\
 & x \in X & \rightarrow \text{core constraints}
\end{array}
$$

Benefits of Slice Modeling

- Reuse program structure and core data
  - Memory and I/O savings
- Use previous solution as starting point for next solve
  - cold/warm/hot starts
- Efficiency in solution allows for larger and more complex problems to be processed

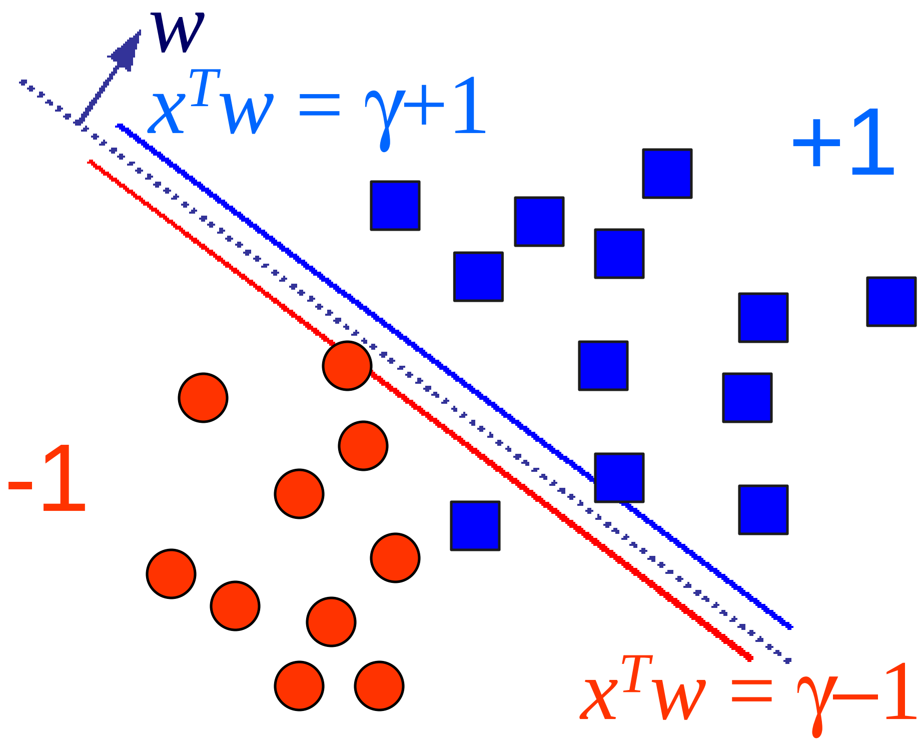
DEA Models as Slice Models

Applying slicing to Data Envelopment Analysis problems:

k	GAMS/CPLEX Time	GAMS/DEA Time
60	4.63 sec	0.43 sec
104	7.02 sec	0.51 sec
4888	16 hours	717.01 sec

Example: Breast Cancer Diagnosis

- Classify patients’ tumors as malignant (-1) or benign (+1)



Example, continued

- Train on data for which outcome is known
- Use cross-validation to test classifier
- Model with a linear classifier and a support vector machine

k-fold Cross-Validation

- Divide data into *k* subsets of approximately equal size
- Train *k* times, leaving out one subset each time
- Use the omitted subset as testing data to estimate error

Support Vector Machine Model (Quadratic Program)

Look for a classifier with the widest margin between classes ($2/\|w\|$) and which minimizes misclassification (z):

$$\begin{aligned} \min_{w,\gamma,\, z} \quad & \tfrac{1}{2}\|w\|^2 + Ce^T z \\ \text{subject to} \quad & D^k(A^k w - \gamma) + z \geq 1 \\ & z \geq 0 \end{aligned}$$

Linear Classification

Replace Euclidean norm measurement with sup-norm measurement:

$$\begin{aligned} \min_{w,\gamma,\, z} \quad & \|w\|_1 + Ce^T z \\ \text{subject to} \quad & D^k(A^k w - \gamma) + z \geq 1 \\ & z \geq 0 \end{aligned}$$

Linear Classification as Linear Program

Convert to linear program by adding new variables y:

$$\begin{aligned} \min_{w,\gamma,\, z,\, y} \quad & e^T y + Ce^T z \\ \text{subject to} \quad & D^k(A^k w - \gamma) + z \geq 1 \\ & y \geq w \geq -y \\ & z \geq 0 \end{aligned}$$

Slice Models in GAMS

```
loop(k,
  slice(k) = yes;
  solve model;
  slice(k) = no);
```

- Solve each problem independently
- Regenerate model
- Repeat core data
- No restarting capability

GAMS/DEA Interface

- Remove loop from model; identify slice constraints by key word `slice`
- Pass all data to interface at once
- Interface defines individual problems automatically
- CPLEX is underlying solver

Deletion Slice Model

Cross-validation models can be formulated as *deletion* slice models:

- Define specific problems by what gets *deleted*
- Slice set is the fold set
- Only specify *testing sets* for each fold
- Declare equations over testing sets

GAMS/DEA Equations (Linear Model)

```
alias(k,slice);
equations obj_def, sep_def(p,slice);


obj_def..
obj =e= sum(f,y(f)) + C*sum(p, z(p));


sep_def(p,k)$test(p,k)..
D(p)*(sum(f, A(p,f)*w(f)) - gamma) + z(p) =g= 1;
```

GAMS/DEA Equations (Linear Model)

```
alias(k,slice);

equations sep_def(p,slice);

sep_def(p,k)$test(p,k)..
D(p)*(sum(f, A(p,f)*w(f)) - gamma) + z(p) =g= 1;
```

GAMS/DEA Solve (Linear Model)

```
option lp = dea;  train.optfile = 1;

*Generate testing sets
loop(k,
$batinclude gentestset “p,k"
);

solve train using lp minimizing obj;
```

GAMS/DEAQP: Quadratic Model

- Extend GAMS/QPWRAP to slice models
- Formulate QP model with
  - linear GAMS model
  - Q matrix specified in text file
- GAMS/QPWRAP passes Q matrix to solver
- Q must be part of core data

Testing Models

- Only one call to the solver → only last solution available to GAMS
- Solutions for all problems written to dea.sol
- Testing independent of training
- Write out solutions (`w, gamma`) to dea.sol for later testing

Results on WDBC Database

Perform all-but-one testing:

Model	GAMS training time	GAMS/DEA training time	Testing accuracy
LP	541.71 sec	269.03 sec	95.08%
QP-CPLEX	7403.03 sec	7221.74 sec	95.42%
QP-CONOPT	519.3 sec	*343.18 sec*	
QP-PATH	691.9 sec	*245.45 sec*	

Conclusions and Future Work

- Model distinct from solver: can easily change problem type
- Model easy to modify
- Use state-of-the-art solvers
- Extend slice techniques to other domains
  - Change solvers without rewriting model
  - Test other formulations and algorithms