# High-Dimensional Data Inference for Automobile Insurance Premia Estimation

Yoshua Bengio, Nicolas Chapados,
Charles Dugas, Joumana Ghosn,
Ichiro Takeuchi, Pascal Vincent

Laboratoire d'informatique des
systèmes adaptatifs
Université de Montréal
www.iro.umontreal.ca/∼lisa

May 2001

# Overview

- Setting the Problem
- Classical Models from Actuarial Practice
- Models from Statistical Learning Algorithms
- Experimental Setting
- Experimental Results
  - Comparing Predictive Accuracy
  - Comparing Premia Distribution
  - Evaluating Fairness
- Conclusions and Future Prospects

# 1. The problem

- **What:** Estimate the premia for customers of automobile insurance (for 1 year)
- Given **input profile** $x$ (representing a customer), estimate expected incurred amount

$$p(x) = E[A|x].$$

- **Critera:** Maximize precision and ensure fairness (mathematically, minimize **squared error**).

## 2a. Statistical Difficulties

- Most drivers don't have any accident in any given year:
  - **But** information about premia is given by those customers who do make a claim;
  - **Hence**, must learn from the **tails of the distribution** (**"outliers"**) — significant for the poor performance of SVM's on this task.
- Large number of variables, with nonlinear interactions.

## 2b. Computational Difficulty

- Extremely large number of training examples — $O(10^7)$; models are hard to train to good minima.

## 3a. Classical Method: GLM

- **Generalized Linear Models** have the overall general expression:

$$p(x) = \exp\left(\beta_0 + \sum_{i=1}^{n} \beta_i x_i\right)$$

- Exponentiation ensures that premia are **positive**.

## 3b. Classical Method: Table-Based

- Typically, tables constructed by hand giving a premium as a function of a few input variables.
- May include "correction terms" to adjust the premia according to various criteria.

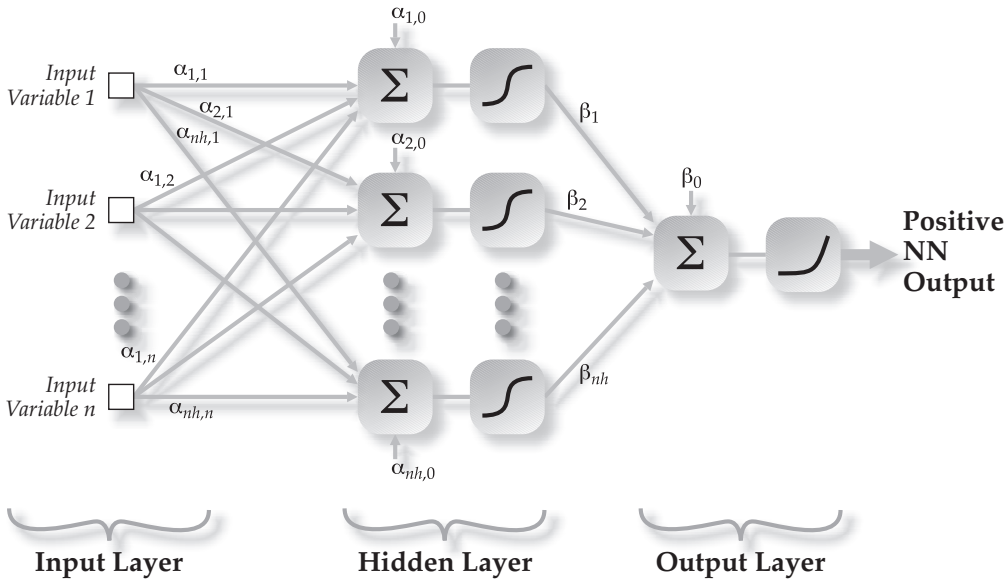
# 4. Positive-Output Neural Networks

- Feed-forward one-hidden-layer neural network whose output is constrained to be positive:

$$p(x) = F\left(\beta_0 + \sum_{i=1}^{n_h} \beta_i \ \tanh\left(\alpha_{i,0} + \sum_{j=1}^{n} \alpha_{i,j} x_j\right)\right)$$

- Softplus is the integral of the sigmoid function:

$$F(y) = \log\left(1 + e^y\right)$$

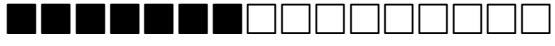


# 5. Mixture of Experts

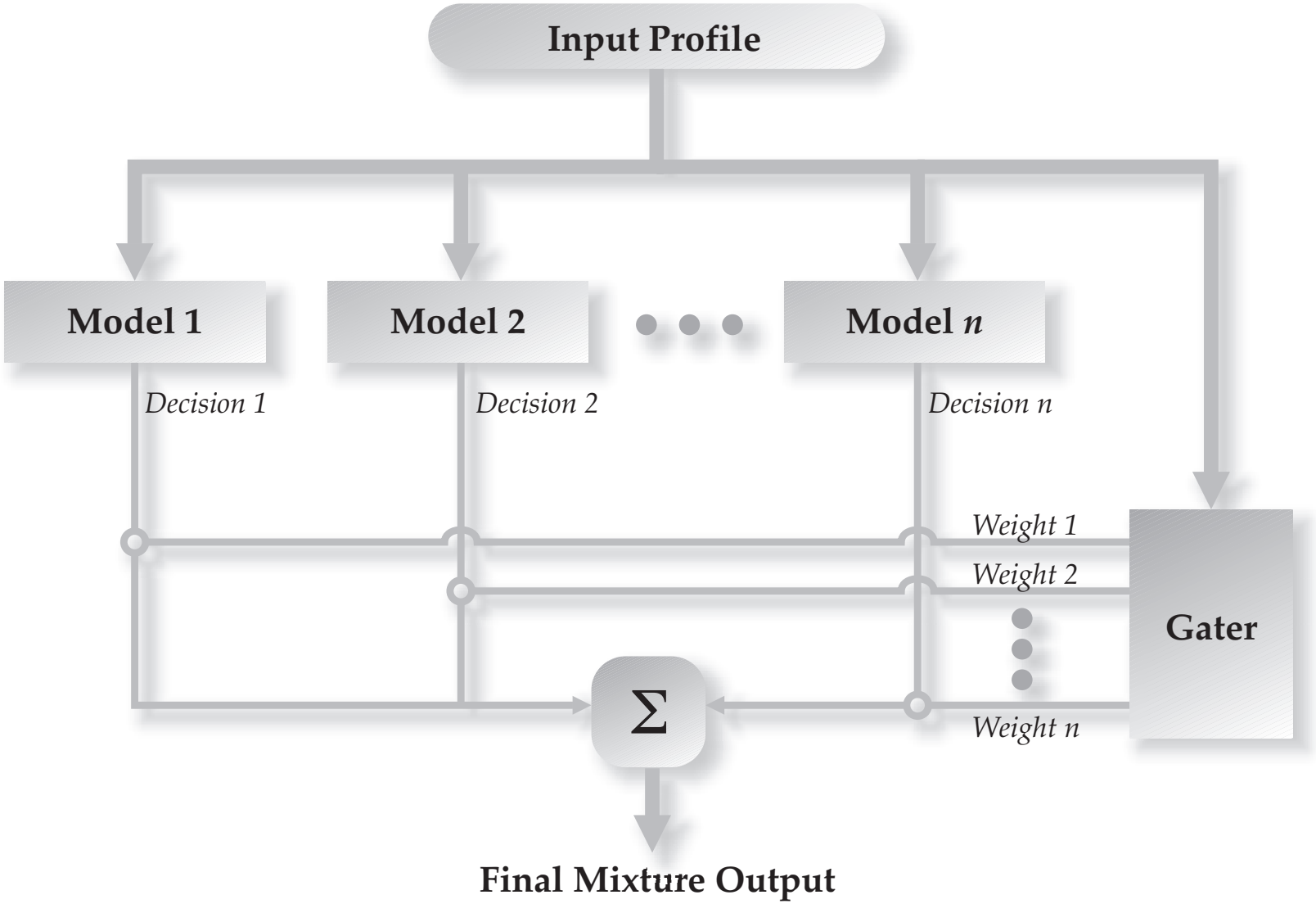
- **Why**: split a hard regression task into **easier subtasks**.
  - Here: separate customers between the groups "large claims", "small claims" and "no claim".
  - Need a **gater model** to give the probability that a given input profile belongs to the the group of large or small claims.
- Mathematically, the prediction is given by

$$p(x) = \sum_{g \in \mathcal{G}} p_g(x)\ P[g|x]\ .$$

  - Probability of each group $P[g|x]$: given by the **gater**.
  - Outputs $p_g(x)$: given by the individual **experts**.
- **Training** is relatively easy, since
  claim amount determines the group $g$.



# 5. Mixture of Experts (continued)



# 6. Issues in Statistical Learning

- **Want** good **generalization** performance
- **Cannot** rely on the **training MSE** to compare models:
  **low training error** $\not\Longrightarrow$ **low generalization error**

# 7. Methodology

# 8. Overview of Results

- After several thousands of experiments:
  - **Best overall model:** mixture model;
  - **Best gater:** feed-forward neural network with softmax output function;
  - **Best expert models:** feed-forward neural network with softplus output function.
- Compare this **Mixture Model** against an existing **Rule-Based Model**.
  - The Rule-Based Model is currently used by a large insurance company to price its contracts;
  - Built by hand: many tables and adjustment rules;
  - Constitutes a **real-world baseline**.

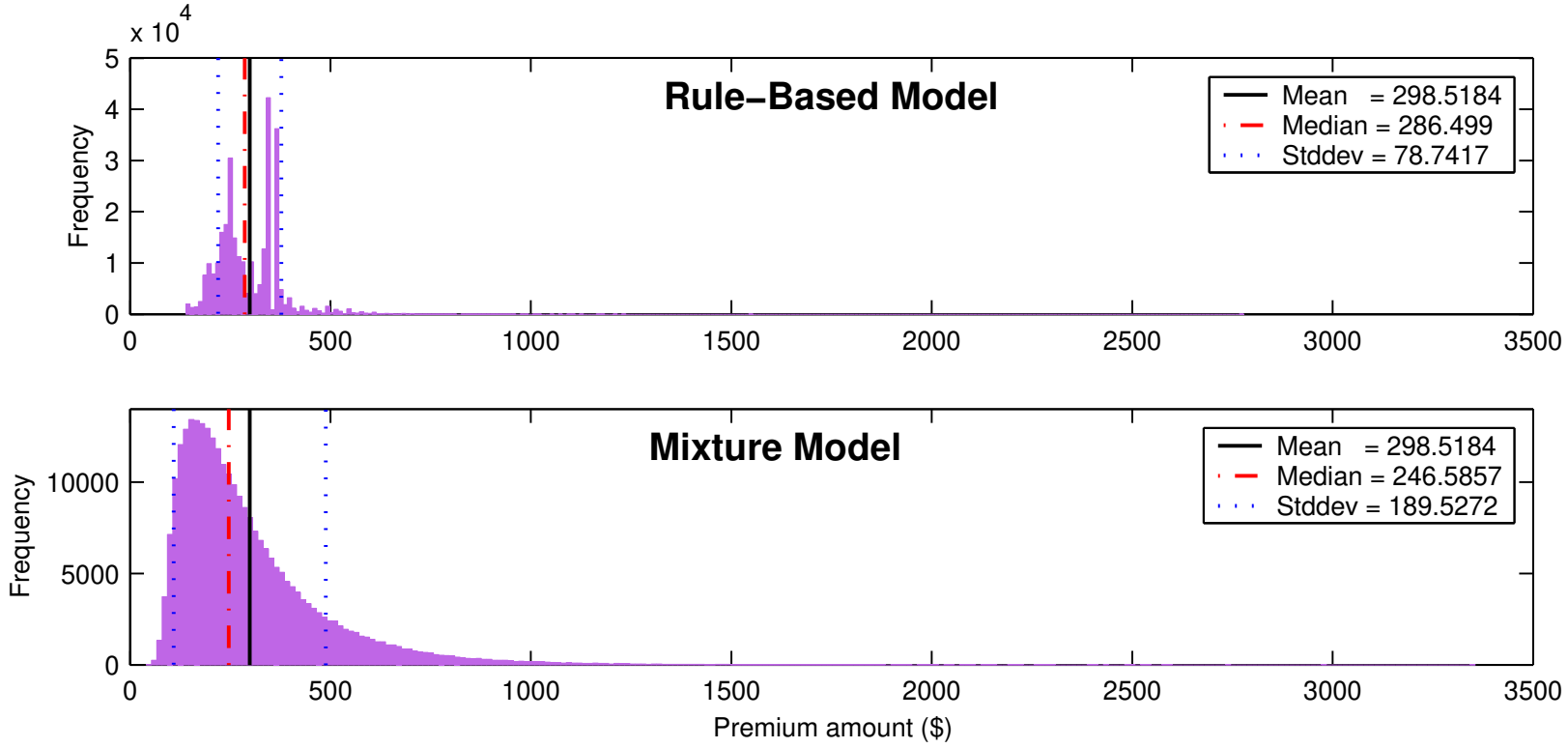
# 9. Comparing Predictive Accuracy

- **Kind of Loss:** tracking category used for actuarial purposes (5 different types).
- **Mixture** yields **significantly lower MSE**.
  - Better predictions

| Kind of Loss | MSE Difference Rule minus Mixture | 95% Confidence Interval Lower | Higher |
|---|---|---|---|
| Group 1 | 20669.53 | ( −4682.83 | 46021.89 ) |
| Group 2 | 1305.57 | ( 1032.76 | 1578.37 ) |
| Group 3 | 244.34 | ( 6.12 | 482.55 ) |
| Group 4 | 1057.51 | ( 623.42 | 1491.60 ) |
| Group 5 | 1324.31 | ( 1077.95 | 1570.67 ) |
| **Total amount** | 60187.60 | ( 7743.96 | 112631.24 ) |

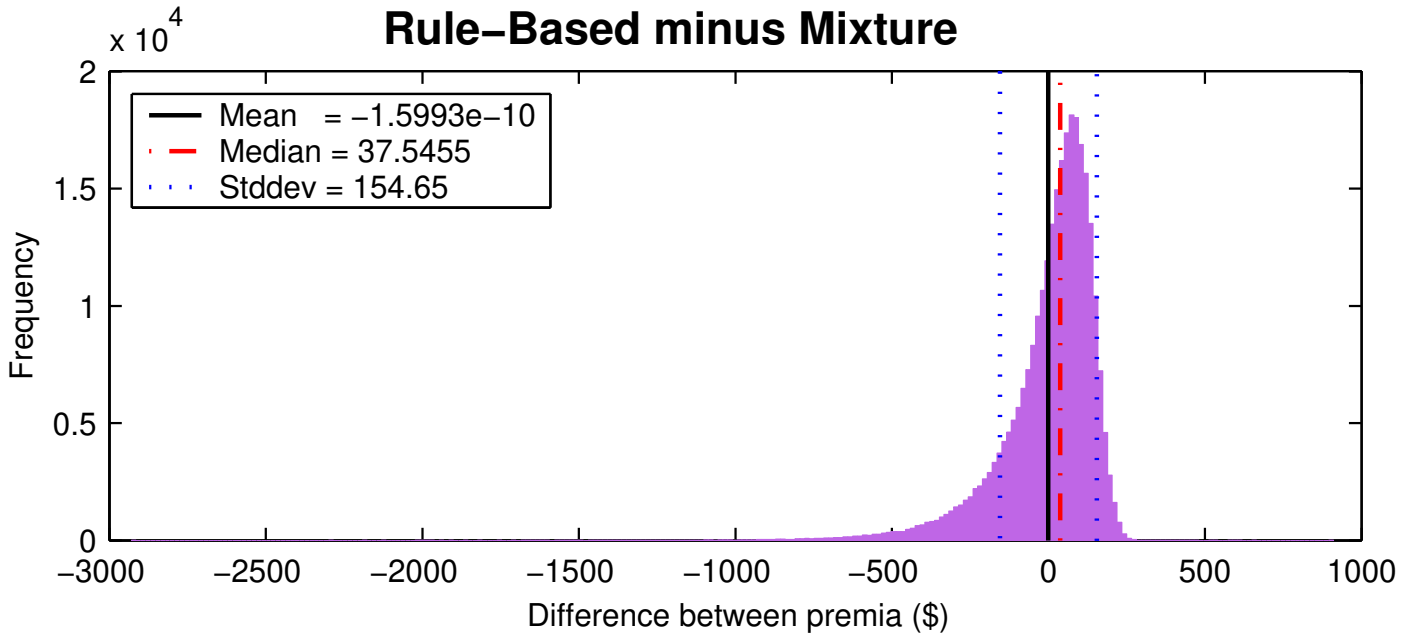
# 10a. Comparing Premia Distributions

- **Mixture model** yields a considerably **smoother** premia distribution (on test-set data), and is more **spread out**.



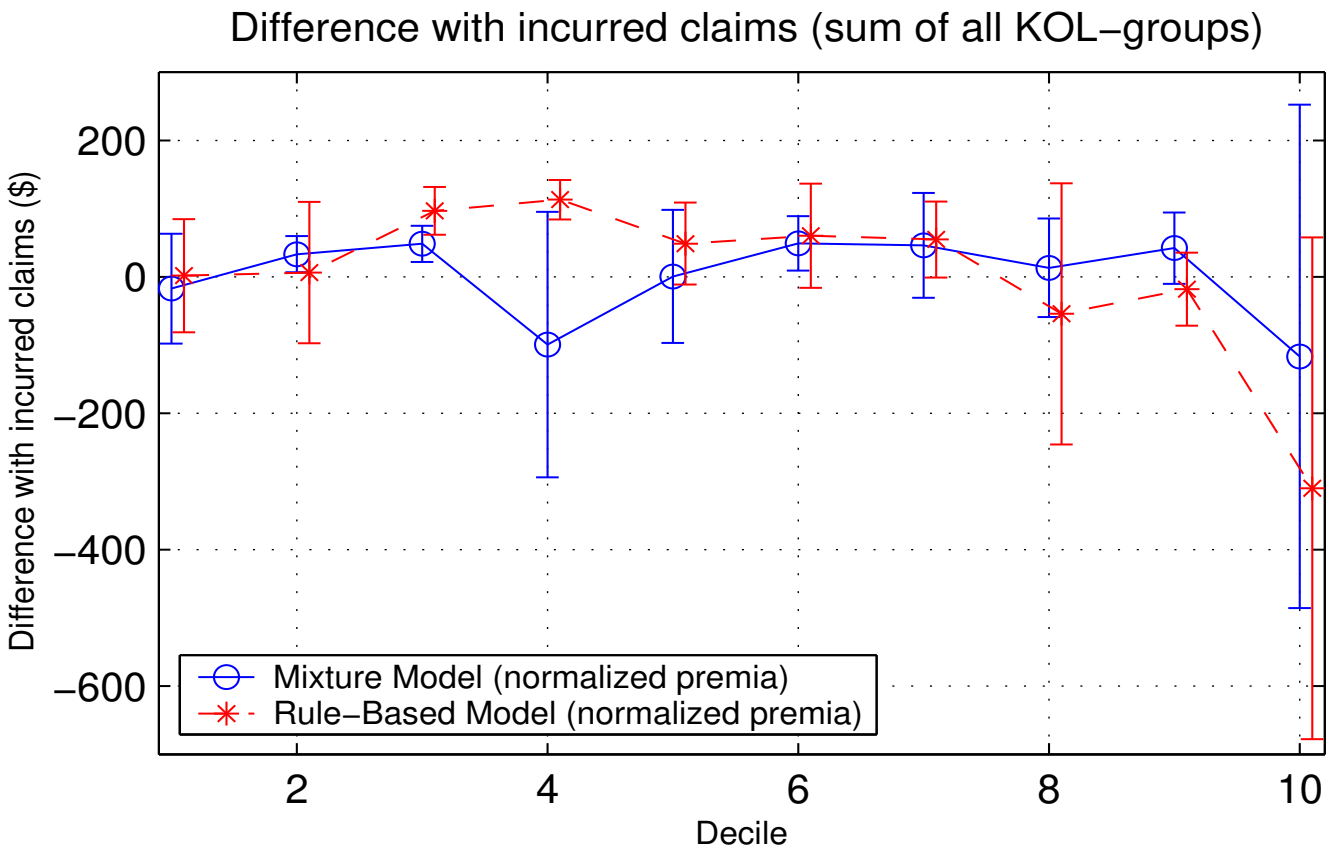
# 10b. Comparing Premia Differences

- Premia difference distribution is **negatively skewed**, but has $> 0$ **median** for a mean of zero.
  - Rule-Based Model **undercharges risky customers**
  - Rule-Based Model **overcharges typical customers**



# 11. Evaluating Fairness

- Compare average incurred amount and premia within each decile of the premia distribution.
- Both models are **generally fair** to subpopulations.



# 12. Conclusions

- **Large opportunity** for using statistical learning algorithms in the insurance industry:
  - Yield premia that better reflect the underlying risk.
- **Using all available input variables** (34 vs. 10) yields considerably more accurate premia:
  - Classical table- and rule-based methods suffer from the **curse of dimensionality**.
- **Tuning the learning architecture** to the problem:
  - Tried Linear models, GLM, CHAID, CHAID+Linear, Regression SVM, ordinary neural networks, etc.
  - Best model involves the **a priori knowledge** that premia should be positive.