Computing for Medicine

Data Science: Lasso Feature Selection, Agentic AI

Tavpritesh Sethi

Approaches for High Dimensional Data

Low rank approximations

- Principal Components Analysis
- Multidimensional scaling
- Factor analysis
- Principal Coordinate Analysis
- Latent Semantic Analysis
- TF-IDF

Machine Learning Approaches

- Filter Based Feature Selection
- Wrapper Based Feature Selection
- Embedded approach
- Bayesian Networks

Metabolic profiles in tuberculosis.

Description

Relative abundances of metabolites from serum samples of three groups of individuals

Details

A data frame with 136 observations on 425 metabolic variables.

Serum samples from three groups of individuals were compared: tuberculin skin test negative (NEG), positive (POS) and clinical tuberculosis (TB).

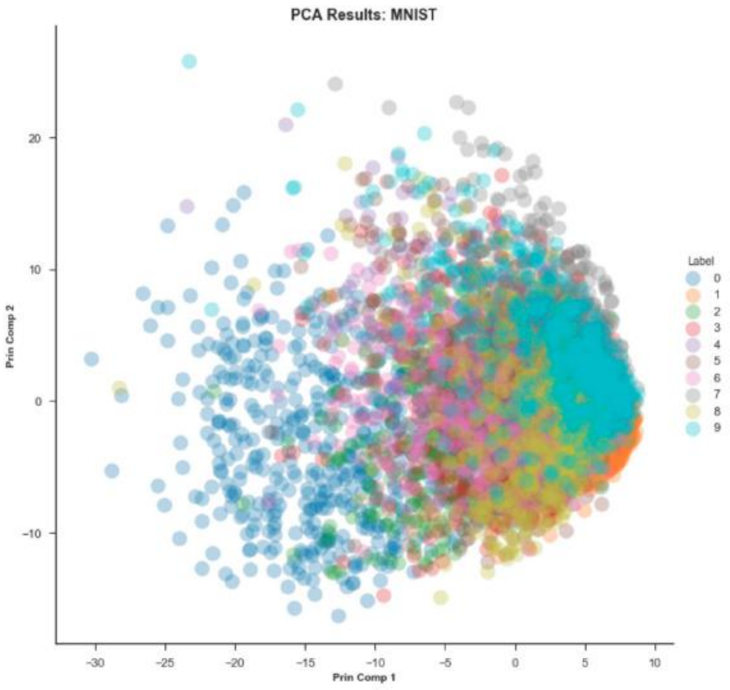
References

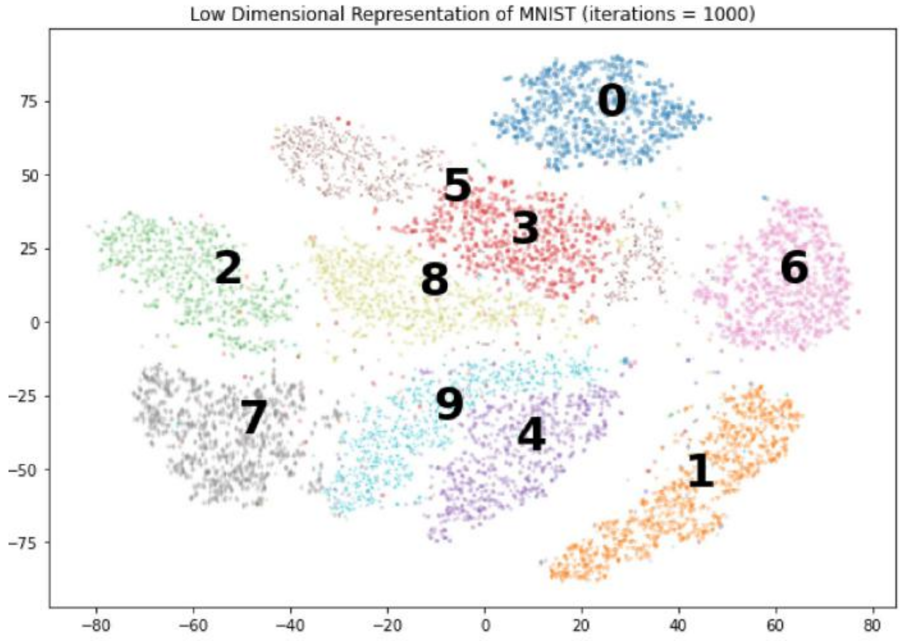
Weiner J 3rd, Parida SK, Maertzdorf J, Black GF, Repsilber D, et al. (2012) Biomarkers of Inflammation, Immunosuppression and Stress Are Revealed by Metabolomic Profiling of Tuberculosis Patients. PLoS ONE 7(7): e40221. doi:10.1371/journal.pone.0040221

Examples

```
data(metabo)
# maybe str(metabo) ; plot(metabo)
```

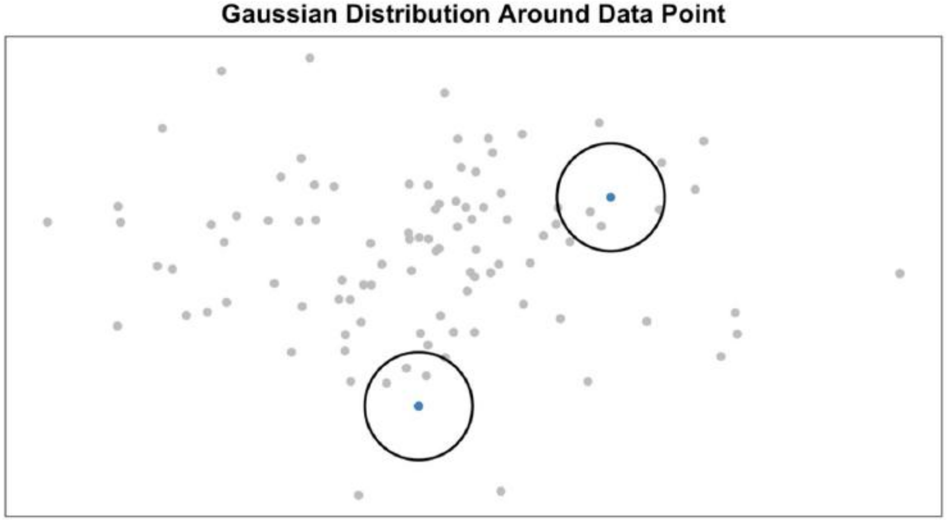
t-Stochastic Neighbor Embedding





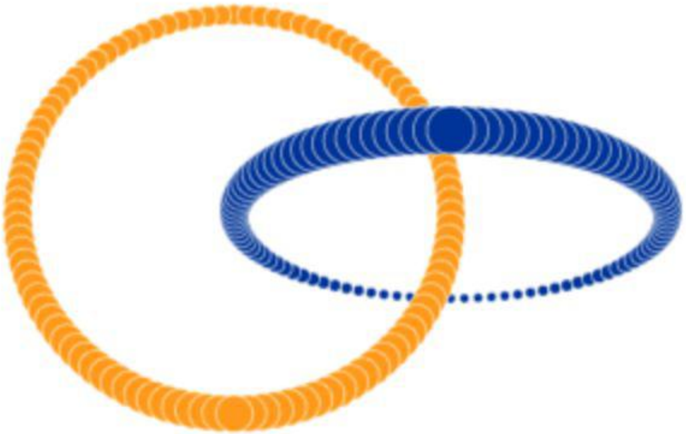
Why t-SNE

- Nonlinear Dimensionality Reduction
- calculates a similarity measure between pairs of instances in the high dimensional space and in the low dimensional space.
- It then tries to optimize these two similarity measures using a cost function.



https://towardsdatascience.com/t-sne-clearly-explained-d84c537f53a

Nonlinearly Separable Data



Why t?

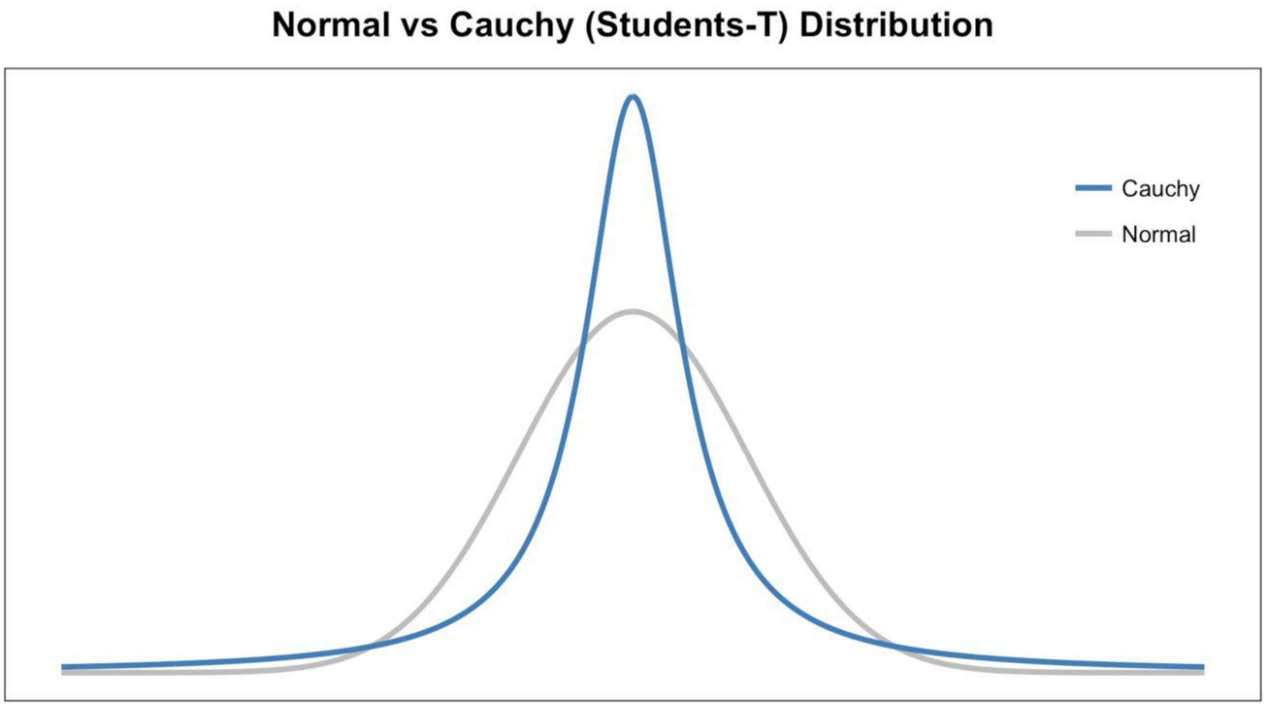
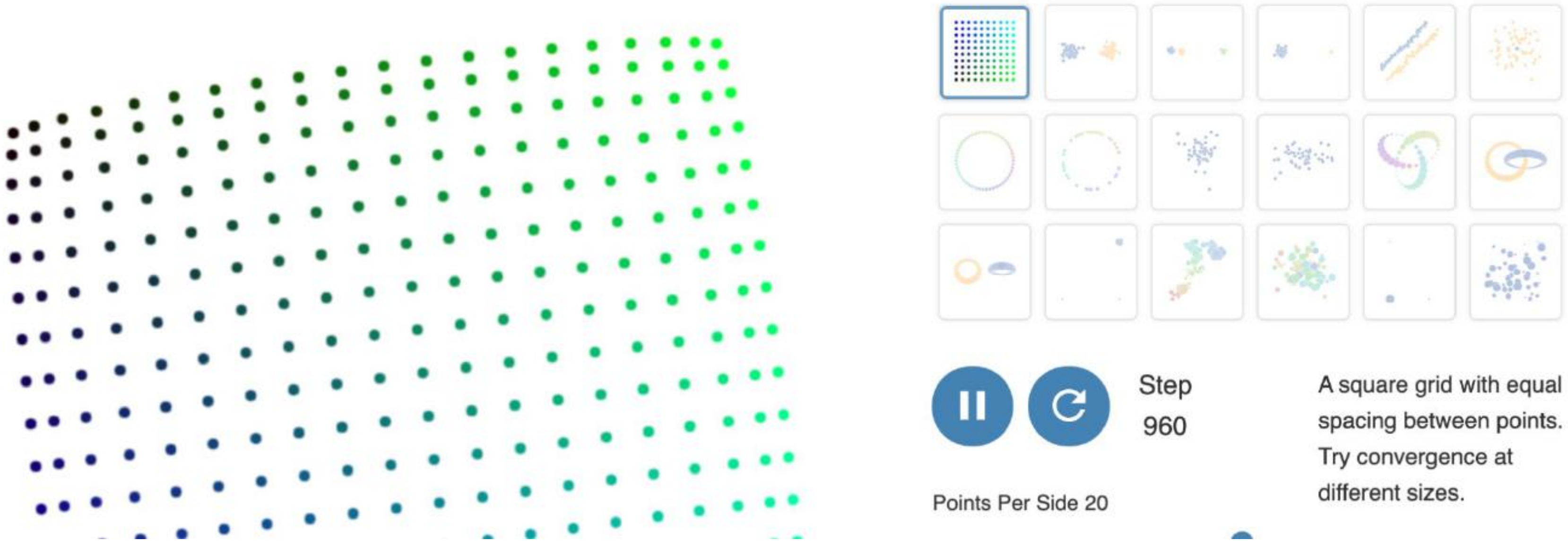


Figure 3 — Normal vs Student t-distribution

The Gaussian distribution or circle can be manipulated using what’s called perplexity, which influences the variance of the distribution (circle size) and essentially the number of nearest neighbors.

How to Use t-SNE Effectively

Although extremely useful for visualizing high-dimensional data, t-SNE plots can sometimes be mysterious or misleading. By exploring how it behaves in simple cases, we can learn to use it more effectively.



https://distill.pub/2016/misread-tsne/

Feature Selection

- Filter Methods eg. correlation matrix using Pearson Correlation).
- Uses the Machine Learning model itself as it’s evaluation criteria (eg. Forward/Backward/Bidirectional/Recursive Feature Elimination).
- Embedded Methods examines the different training iterations of our ML model and then ranks the importance of each feature based on how much each of the features contributed to the ML model training (eg. LASSO Regularization).

https://towardsdatascience.com/feature-selection-techniques-1bfab5fe0784

Filter Approach

- Uses a quality measure that is independent from the classification algorithm
- A selection criterion (or feature relevance measure) is used
- The classifier is built and evaluated only once at the end of the process
- No iteratively building and evaluated in a loop
- Two subtypes
    - Considering each feature as independent of each other
    - Taking correlations between features into account (Correlation Based Feature Selection)

Correlation based Feature Selection

- Principle: A good feature subset is one that contains features highly correlated with the class, yet uncorrelated with each other.

$$Merit_s = \frac{k\overline{r_{cf}}}{\sqrt{k + k(k-1)\overline{r_{ff}}}}$$

- Makes use of optimization techniques to achieve this- genetic algorithms, hill climbing search and beam search

Embedded Approach: LASSO & Ridge Regression

- Stands for Least Absolute Shrinkable and Selection Operator
- Regularization - process of adding information to prevent the over-fitting problem

Lasso Cost Function

$$J(m) = \sum_{i=0}^{n} (\hat{y} - y_i)^2 + \lambda \, |slope|$$

Ridge Cost Function

$$J(m) = \sum_{i=0}^{n} (\hat{y} - y_i)^2 + \lambda \, (slope)^2$$

Quiz: Which one will you use for feature selection?

Regularization

- p> N (sparse situation) - minimize a modified objective function
- A regularization term added to prevent the coefficients from being too large
- λ times the L1-norm - Lasso
- λ times the L2-norm - Ridge
- Lasso: at a certain λ, all coefficients will eventually become 0

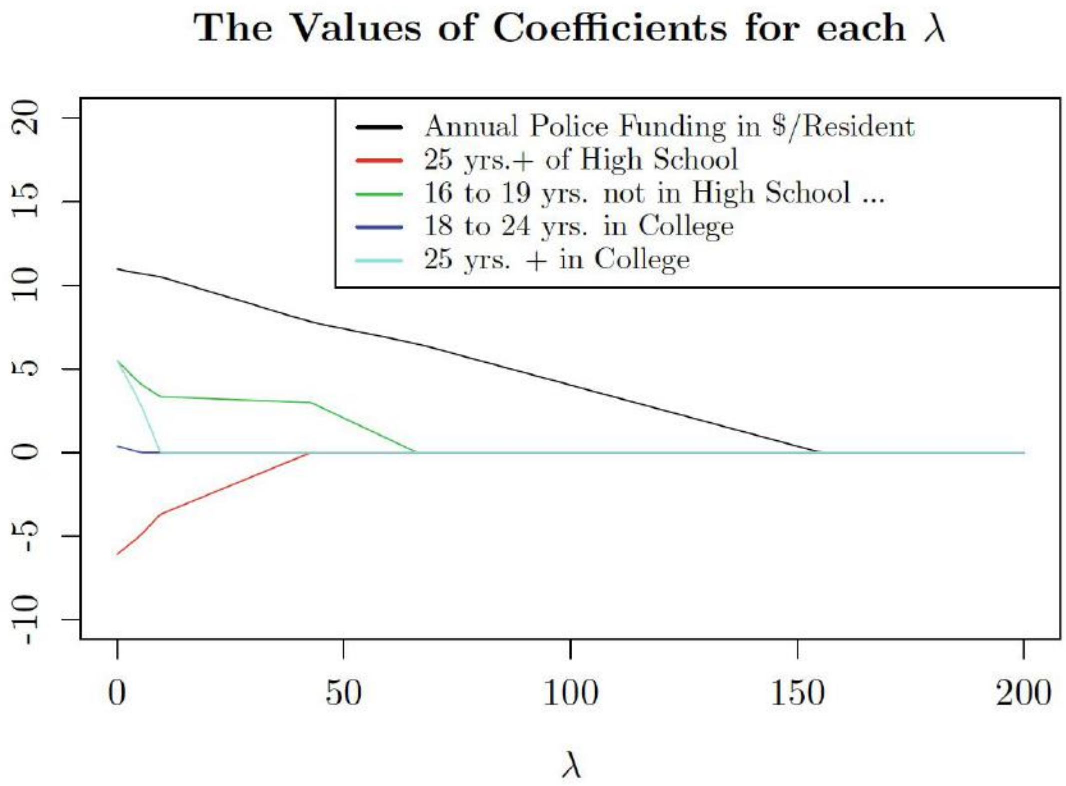
$$L := \frac{1}{2N} \|y - X\beta\|^2 + \lambda \|\beta\|_1$$

Example: US Crime Data (Hastie)

https://web.stanford.edu/~hastie/StatLearnSparsity/data.html

Column	Cov./Res.	Definition of Variable
1	response	crime rate per 1 million residents
2		(we currently do not use this)
3	covariate	annual police funding
4	covariate	% of people 25 years+ with 4 yrs. of high school education
5	covariate	% of 16–19-year-old persons not in high school and not high school graduates
6	covariate	% of 18–24-year-old persons in college
7	covariate	% of people 25 years+ with at least 4 years of college education

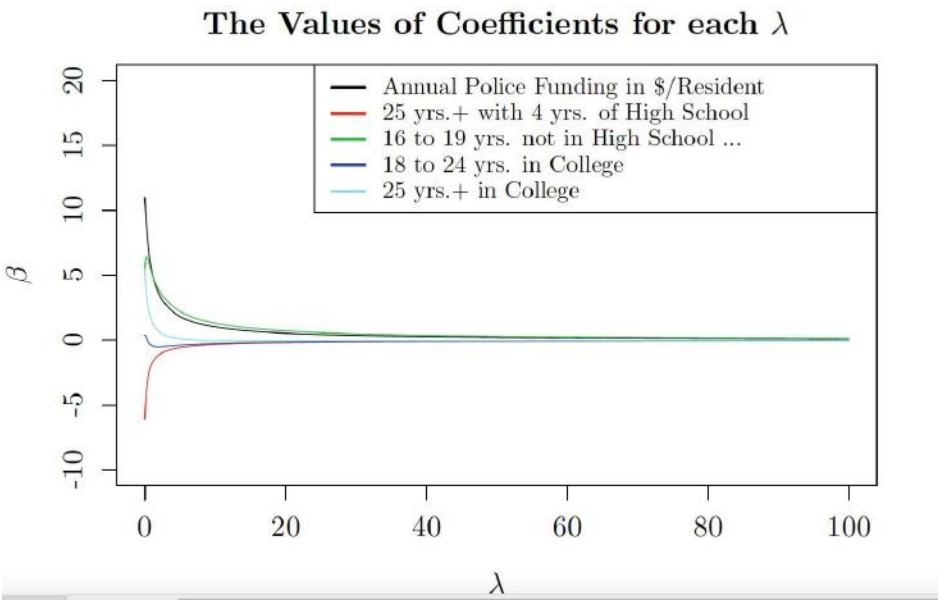
Warm starting the calculation of λ



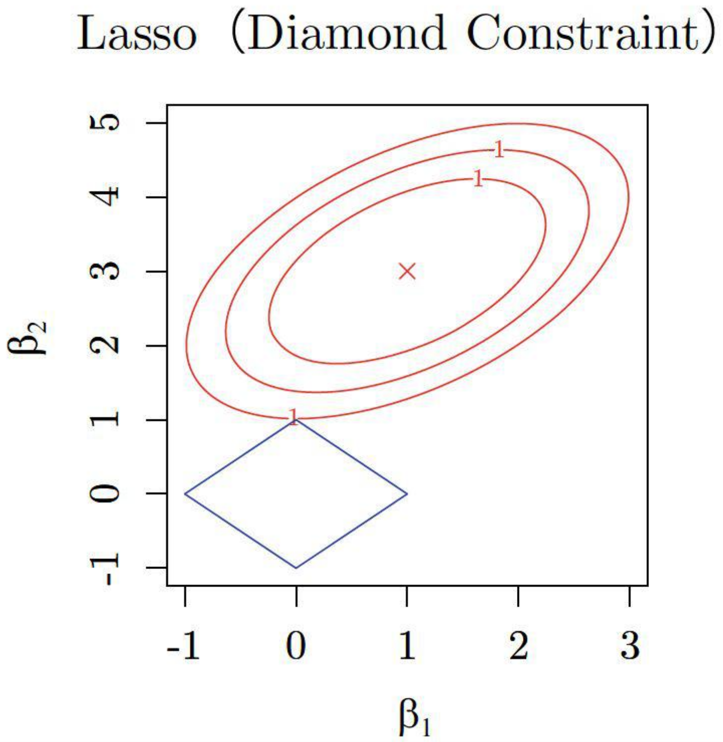
- begin with a λ large enough that every coefficient is zero
- make λ smaller gradually.
- improves the calculation performance by setting the initial value of β to the estimated β from a previous λ

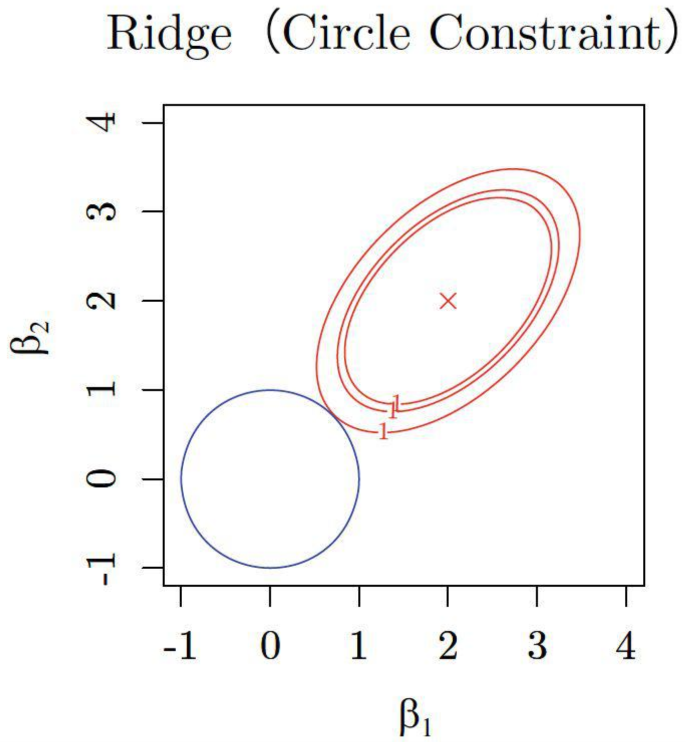
Ridge Regression

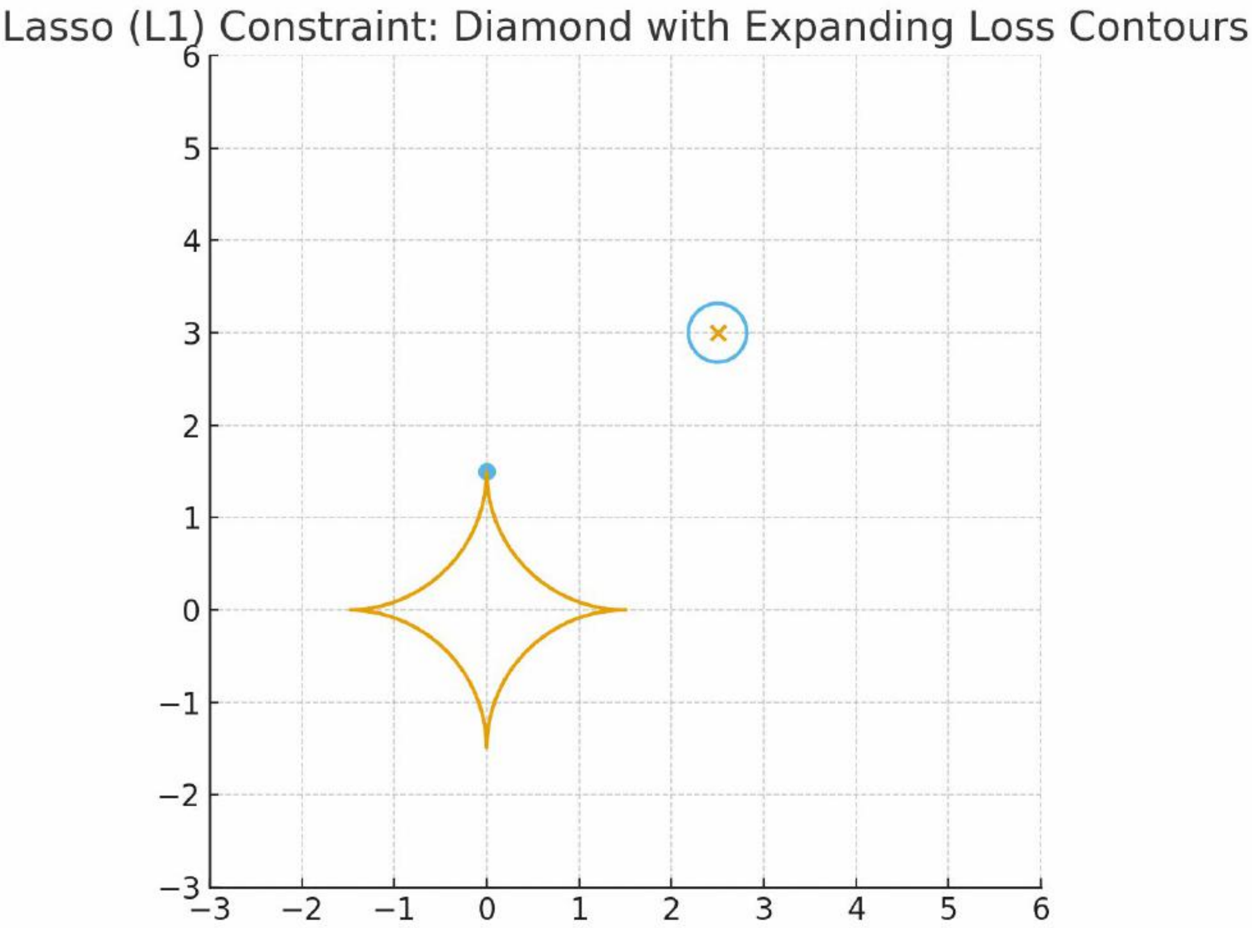
$$L := \frac{1}{N}\|y - X\beta\|^2 + \lambda\|\beta\|_2^2$$



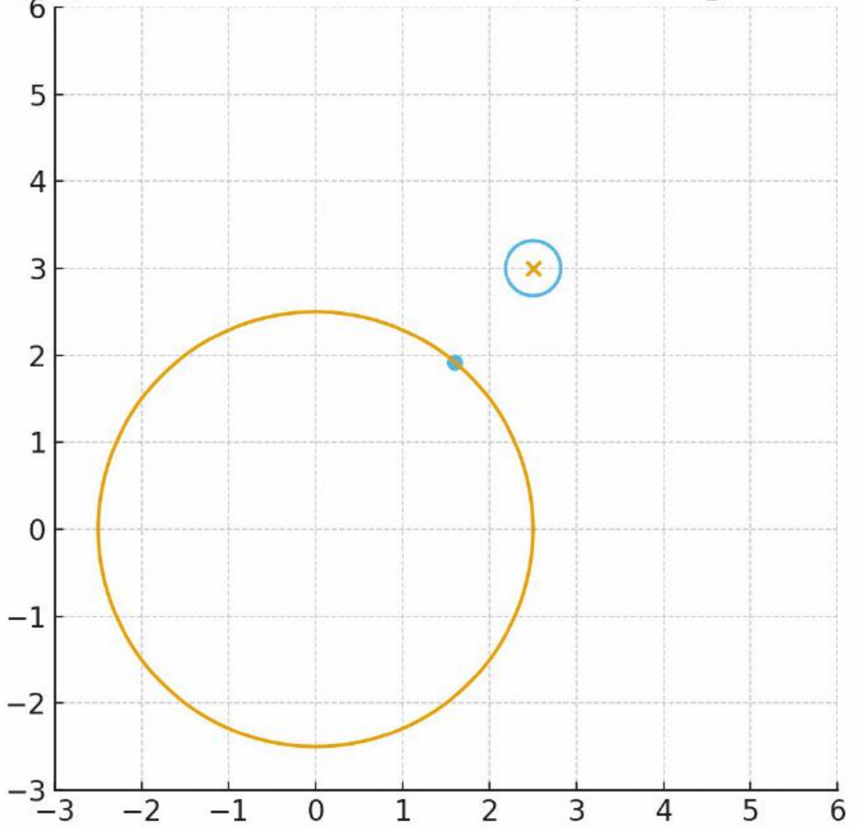
Why does Lasso perform variable selection?



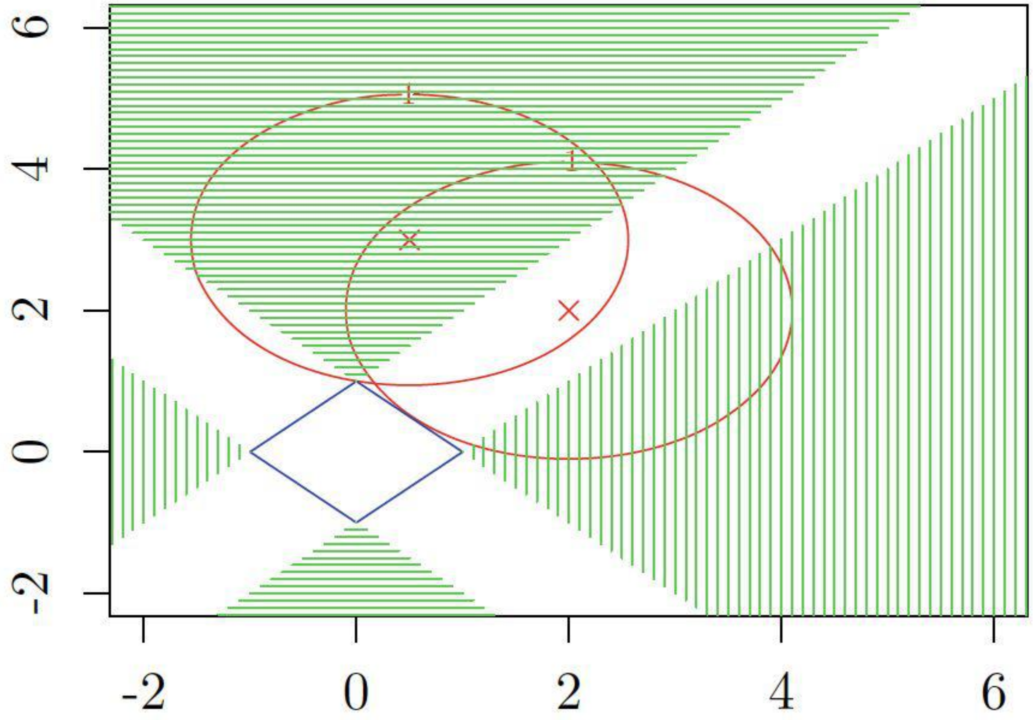



Lasso (L1) Constraint: Diamond with Expanding Loss Contours
6
5
4
3
2
1
0
−1
−2
−3
−3 −2 −1 0 1 2 3 4 5 6

Ridge (L2) Constraint: Circle with Expanding Loss Contours



Further Intuition



However, Ridge regression performs better in explaining if there's collinearity!

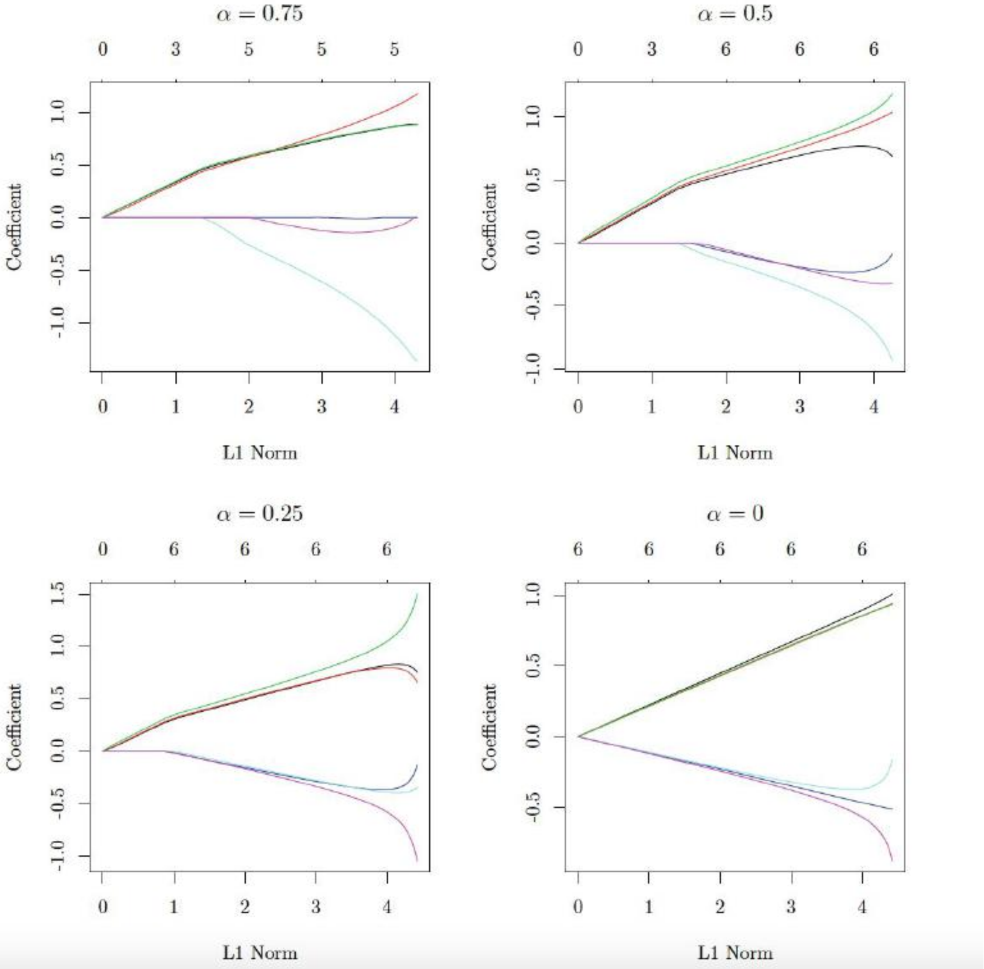
Further variants of Lasso:-

- Group Lasso
- Fused Lasso
- Graphical Lasso

Elastic Net

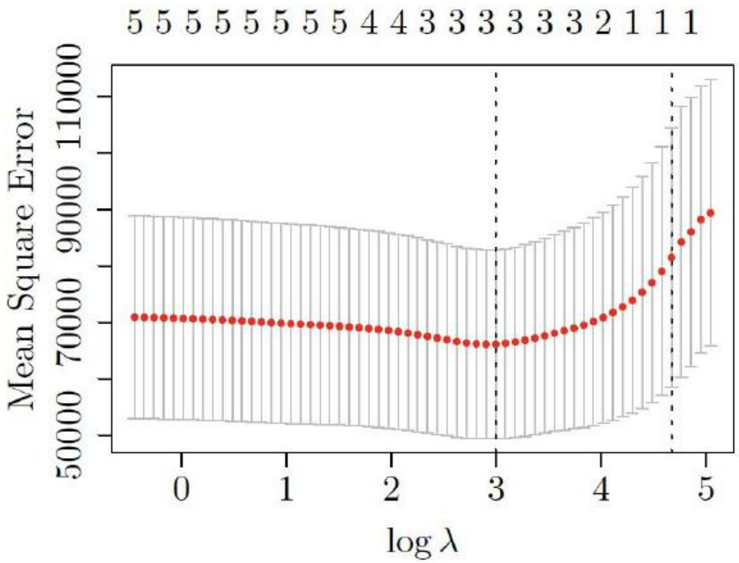
$$L := \frac{1}{2N}\|y - X\beta\|_2^2 + \lambda\left\{\frac{1-\alpha}{2}\|\beta\|_2^2 + \alpha\|\beta\|_1\right\}$$

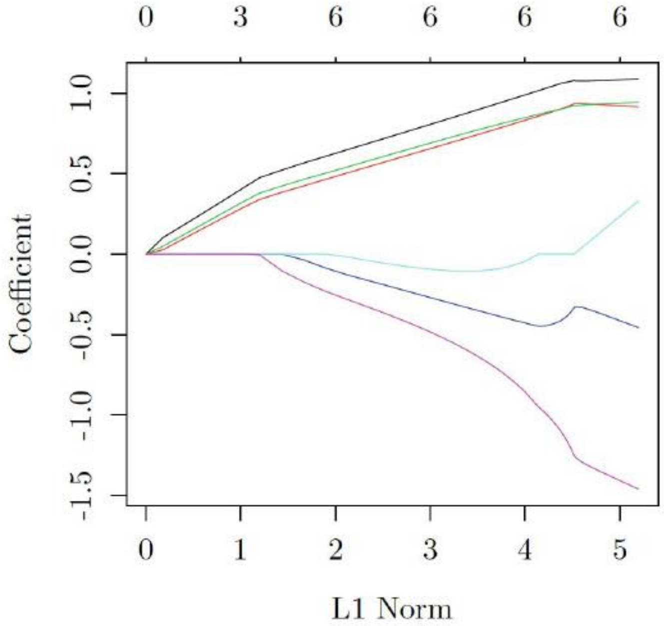
Combines the properties of both Lasso and Ridge



How to set λ

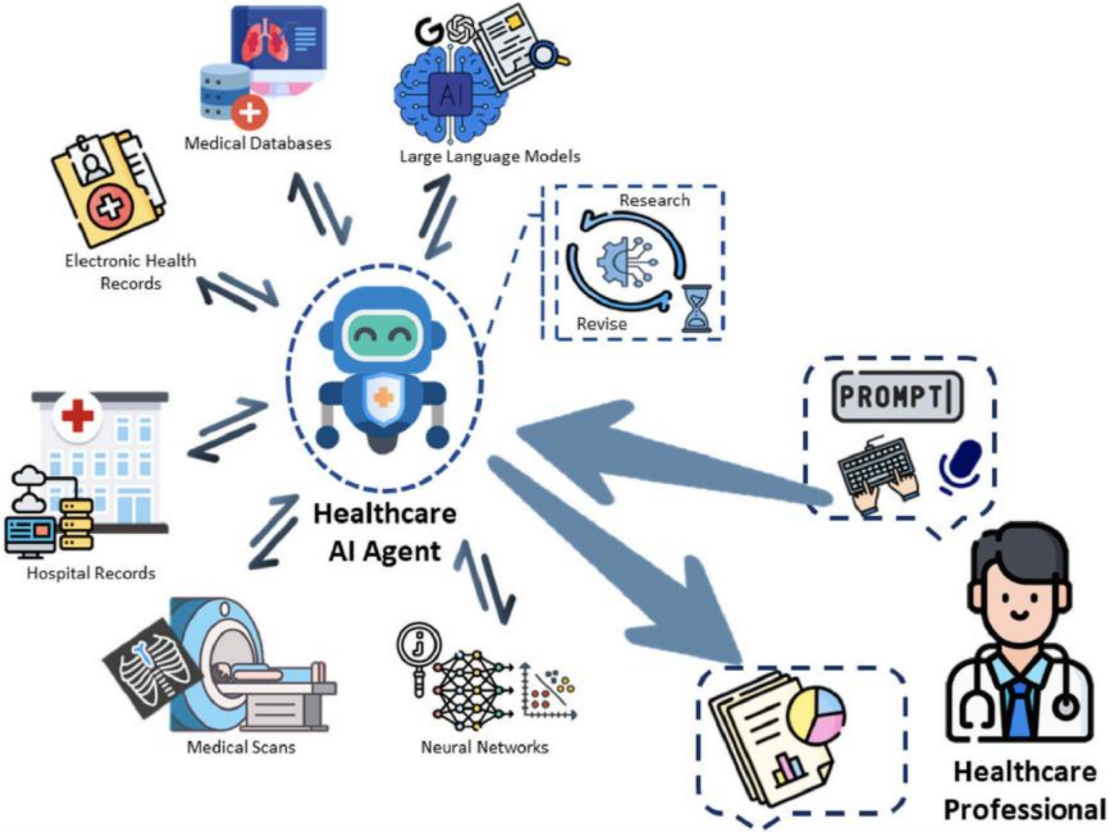
Use n-fold cross-validation to find the optimum λ





Agentic AI in Healthcare

- **Healthcare AI Agent as a central hub** for clinical decision support
- **Ingests multimodal data:** EHRs, hospital records, medical databases, medical scans
- **Core intelligence from AI models:** Large Language Models + Neural Networks
- **Iterative improvement loop:** research → revise → deploy better guidance
- **Clinician-facing interface:** prompt/voice/text input for easy interaction
- **Outputs actionable insights/reports** back to healthcare professionals
- **Goal:** faster, more consistent, data-driven care at the point of use

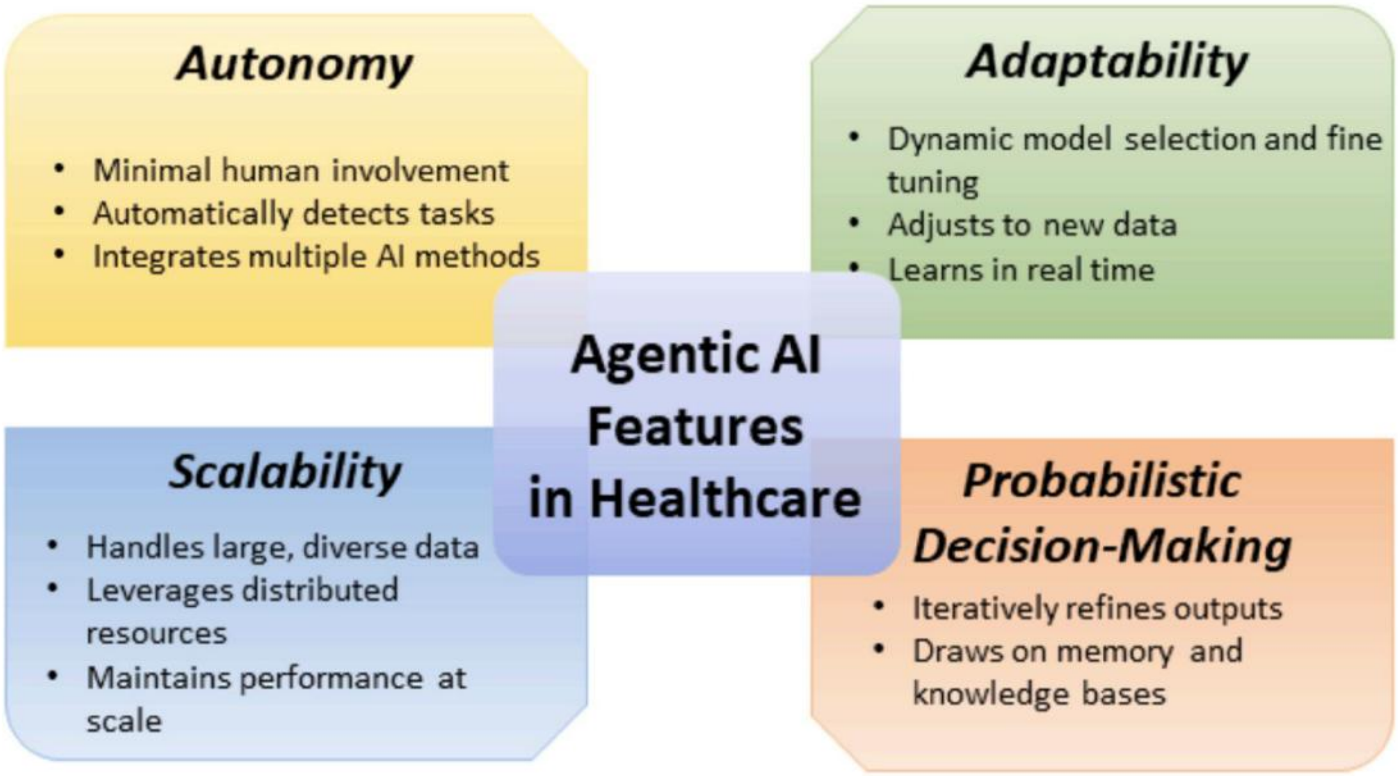


Potential Applications of Agentic AI

Categorization of AI agent types in healthcare with key applications, users, and technologies.

AI agents	Key applications	Healthcare categories	Main users	Key AI technologies
Image base agents	Disease diagnosis, early detection, report generation	Diagnosis, Clinical decision support	Radiologists, Doctors	Computer vision (CNNs, ViTs), MLLMs for image-text integration
predictive analytics agents	Risk prediction, disease progression forecasting, patient outcomes	Clinical Decision Support, Treatment and Patient Care, Drug Discovery & Research	Doctors, Care Teams	Predictive Modeling, including supervised ML, ensemble methods, and time-series analysis
Conversational agents	Symptom checking, patient triage, virtual consultations	Patient Engagement and Monitoring	Patients, General Practitioners	NLP, Dialogue Systems, Pretrained LLMs
NLP agents	Processing clinical notes, summarizing EHRs, extracting insights	Operations and Administration, Clinical Decision Support	Medical Coders, Analysts	NLP, Pretrained LLMs
Rule base agents	Following clinical guidelines, alerting for drug interactions	Clinical Decision Support	Doctors, Pharmacists	Rule-Based Reasoning, leveraging logic programming, expert rules, knowledge graphs
Hybrid agents	Combining imaging, text, video, and predictive analytics for decisions	Clinical Decision Support, Diagnosis, Robot-Assisted Surgery	Doctors, Radiologists, Surgeons	Multimodal Learning
ML agents	Disease classification, anomaly detection, treatment planning	Diagnosis, Treatment and Patient Care, Drug Discovery & Research	Data Scientists, Doctors	ML/DL algorithms, RL
Expert system agents	Emulating clinical expertise for diagnosis and planning	Treatment and Patient Care, Clinical Decision Support, Robot-Assisted Surgery	Specialists, Researchers, Surgeons	Knowledge-based systems, rule-based systems
Recommender agents	Suggesting diagnostic tests, personalized treatments	Treatment and Patient Care, Clinical Decision Support	Doctors, Care Teams	Collaborative filtering, recommendation systems, RL

Key Features of an Agentic AI Application



Key Differences Between Traditional AI Agents and Agentic AI in healthcare.

Aspect	Traditional AI agent	Agentic AI
Reasoning Approach	Use domain specific algorithms (e.g., RL, finite state machines). Example: A rule-based chatbot that answers predefined medical FAQs.	Natural language-based reasoning, primarily leverage LLMs. Example: An AI assistant that engages in open-ended medical Q&A and explains conditions based on patient history.
Domain Flexibility	Optimized for a specific task or environment. Example: Optimized for specific tasks such as structured report generation or rule-based classification.	Adaptable to multiple tasks using zero/few-shot learning or dynamic prompt engineering. Example: Can summarize radiology reports and also assist in diagnostic support.
Decision Structure	Rule-based, with a predefined set of actions and goals. Example: An AI system that follows fixed guidelines for triaging emergency cases.	Capable of self-directed sub-goals and iterative planning (e.g., language-based reasoning models). Example: An AI agent that dynamically adjusts triage decisions based on patient data and evolving symptoms.

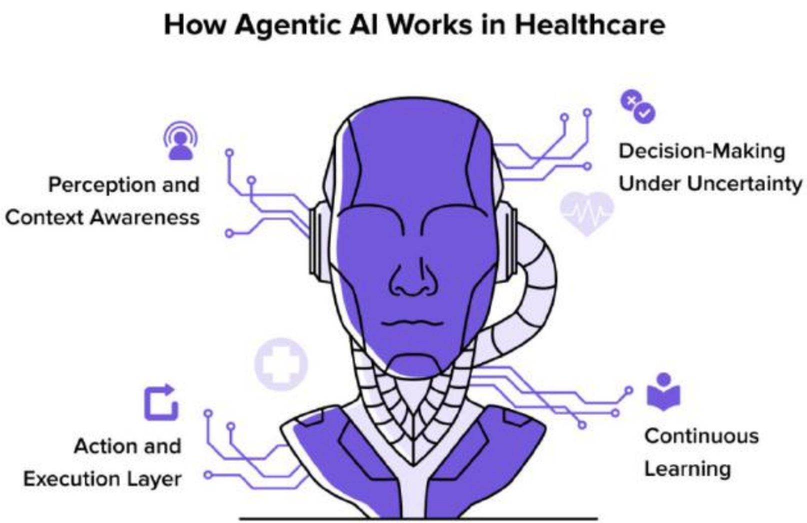
Key Differences Between Traditional AI Agents and Agentic AI in healthcare.

Data & Training	Domain-specific datasets. Example: AI trained only on structured EHR data.	Utilizes large-scale pretraining over vast corpora; can generalize or specialize with minimal updates (few-shot). Example: AI trained on diverse medical texts, adapting quickly to new diseases and guidelines.
Tool Integration	Limited. Example: A standalone AI tool for detecting anomalies in CT scans.	Dynamically calls external APIs, knowledge bases, ML, DL, MLLMs. Example: An AI system that integrates with hospital systems, lab results, and medical literature to provide comprehensive diagnostic insights.

Key Differences Between Traditional AI Agents and Agentic AI in healthcare.

Explanations	Transparent in symbolic systems; moderate interpretability in RL. Example: A decision tree model explaining why a drug is recommended.	Largely blackbox, though techniques for interpretability are emerging. Example: Agentic AI system suggesting treatments based on DL patterns, with efforts to explain its reasoning using attention maps or counterfactual analysis.
Primary Use Cases	Robotics, rule-based decision-making, industrial control systems. Example: A robotic system following predefined surgical procedures.	Unlimited Example: AI-driven virtual assistants for patient monitoring, adaptive treatment planning, personalized medicine, and robotic surgery and many more.

Agentic AI in Healthcare



Use Case	Description
CDSS	Data-Driven Decisions
Diagnosis	Real-time Insights for Accurate Detection
Treatment and Patient Care	Personalized, Adaptive Care Models
Patient Engagement and Monitoring	Empowering Patients with Continuous Feedback
Healthcare Administration and Operations	Streamlining Hospital Management
Drug Discovery and Research	Accelerating the Development of New Medications
Robot-Assisted Surgery	Precision at the Cutting Edge
Chronic Care Agents	Continuous Support for Long-Term Conditions
Surgical Planning	Enhancing Precision Before the Procedure

appinventiv

https://appinventiv.com/blog/agentic-ai-in-healthcare/#:~:text=Use%20Case:%20AI%20in%20Remote,decisions%20in%20managing%20their%20health.