Phase transitions in Generalized Linear Models

Cargèse summer school

August 22, 2018

Léo Miolane

with Jean Barbier, Florent Krzakala, Nicolas Macris & Lenka Zdeborová
Generalized Linear Models

Definition

Statistical model

- One observes for $1 \leq \mu \leq m$

$$Y_\mu \sim P_{\text{out}}(\cdot \mid \langle \Phi_\mu, X^* \rangle)$$

- $X^* \in \mathbb{R}^n$: signal vector of dimension $n$.
- $\Phi_1, \ldots, \Phi_m \in \mathbb{R}^n$: measurement vectors.
- $P_{\text{out}}$: transition kernel.

Goal: recover $X^*$ from $Y$ (and $\Phi$).
Generalized Linear Models

Definition

Statistical model

- One observes for $1 \leq \mu \leq m$

$$Y_\mu \sim P_{\text{out}}(\cdot \mid \langle \Phi_\mu, X^* \rangle)$$

- $X^* \in \mathbb{R}^n$: signal vector of dimension $n$.
- $\Phi_1, \ldots, \Phi_m \in \mathbb{R}^n$: measurement vectors.
- $P_{\text{out}}$: transition kernel.

**Goal:** recover $X^*$ from $Y$ (and $\Phi$).

- When is it information-theoretically possible?
- When is it computationally tractable?
Examples

Some interesting particular cases

» Linear model:

\[ Y = \Phi X^* \]

» Phase retrieval:

\[ Y = |\Phi X^*| \]

» 1-bit CS (“Planted” perceptron):

\[ Y = \text{sign}(\Phi X^*) \]
Examples
Some interesting particular cases

▶ Linear model:

\[ Y = \Phi X^* + \text{Noise} \]

▶ Phase retrieval:

\[ Y = |\Phi X^*| + \text{Noise} \]

▶ 1-bit CS ("Planted" perceptron):

\[ Y = \text{sign}(\Phi X^* + \text{Noise}) \]
Examples
Some interesting particular cases

▶ Linear model:

\[ Y = \Phi X^* + \text{Noise} \]

▶ Phase retrieval:

\[ Y = |\Phi X^*| + \text{Noise} \]

▶ 1-bit CS ("Planted" perceptron):

\[ Y = \text{sign}(\Phi X^* + \text{Noise}) \]

▶ Logistic model:

\[
Y_\mu = \begin{cases} 
+1 & \text{with probability } \frac{1}{1 + \exp(-\lambda \langle \Phi_\mu, X^* \rangle)} \\
-1 & \text{with probability } \frac{1}{1 + \exp(\lambda \langle \Phi_\mu, X^* \rangle)} 
\end{cases}
\]
Setting

Assumptions

\[ \mathbf{Y} \sim P_{\text{out}} \left( \cdot \mid \Phi \mathbf{X}^* \right) \]

- Asymptotic regime: \( n \to \infty, \quad m/n \to \alpha > 0. \)
- \( \mathbf{X}^* = (X^*_1, \ldots, X^*_n)^{\text{i.i.d.}} \sim P_0, \quad \mathbb{E}_{P_0} X^2 = \rho. \)
Setting

Assumptions

\[ Y \sim P_{\text{out}}(\cdot \mid \Phi X^*) \]

- **Asymptotic regime:** \( n \to \infty, \quad m/n \to \alpha > 0 \).

- \( X^* = (X_1^*, \ldots, X_n^*) \text{i.i.d.} \sim P_0, \quad \mathbb{E}_{P_0}X^2 = \rho \).

- \( (\Phi_{i,j}) \) are independent, \( \begin{cases} \mathbb{E}\Phi_{i,j} = 0 \\ \mathbb{E}\Phi_{i,j}^2 = 1/n \\ \sup_{i,j} \mathbb{E}|\Phi_{i,j}|^3 \text{ remains bounded} \end{cases} \)

- \( \mathbb{E}[|Y_\mu|^{2+\epsilon}] \) remains bounded, for some \( \epsilon > 0 \).
Setting

Assumptions

\[ Y \sim P_{\text{out}}(\cdot \mid \Phi X^*) \]

- **Asymptotic regime**: \( n \to \infty, \quad m/n \to \alpha > 0 \).
- \( X^* = (X_1^*, \ldots, X_n^*) \overset{\text{i.i.d.}}{\sim} P_0, \quad \mathbb{E}_{P_0} X^2 = \rho \).
- \((\Phi_{i,j})\) are independent, \( \begin{cases} \mathbb{E} \Phi_{i,j} = 0 \\ \mathbb{E} \Phi_{i,j}^2 = 1/n \\ \sup_{i,j} \mathbb{E} |\Phi_{i,j}|^3 \text{ remains bounded.} \end{cases} \)
- \( \mathbb{E} [ |Y_\mu|^{2+\epsilon} ] \) remains bounded, for some \( \epsilon > 0 \).
- \( x \in \mathbb{R} \mapsto P_{\text{out}}(\cdot \mid x) \) is continuous almost everywhere.
Setting

Assumptions

\[ \mathbf{Y} \sim P_{\text{out}} \left( \cdot \mid \Phi \mathbf{X}^* \right) \]

- **Asymptotic regime:** \( n \to \infty, \quad m/n \to \alpha > 0. \)
- \( \mathbf{X}^* = (X_1^*, \ldots, X_n^*) \overset{\text{i.i.d.}}{\sim} P_0, \quad \mathbb{E}_{P_0} X^2 = \rho. \)
- \( (\Phi_{i,j}) \) are independent,
  \[
  \begin{cases}
  \mathbb{E} \Phi_{i,j} = 0 \\
  \mathbb{E} \Phi_{i,j}^2 = 1/n \\
  \sup_{i,j} \mathbb{E} |\Phi_{i,j}|^3 \text{ remains bounded.}
  \end{cases}
  \]
- \( \mathbb{E} \left[ |Y_{\mu}|^{2+\epsilon} \right] \) remains bounded, for some \( \epsilon > 0. \)
- \( x \in \mathbb{R} \mapsto P_{\text{out}}(\cdot \mid x) \) is continuous almost everywhere.
- \( P_{\text{out}} \) has to be "regularized" by some (small) Gaussian noise:
  \[ \forall x \in \mathbb{R}, \quad "P_{\text{out}}(\cdot \mid x) = \tilde{P}_{\text{out}}(\cdot \mid x) + \mathcal{N}(0, \sigma^2)" \], where \( \sigma > 0. \)
- If \( P_{\text{out}} \) takes values in \( \mathbb{N} \), no need for regularization (\( \sigma = 0 \)).

The statistician knows the model, i.e. \( P_0 \) and \( P_{\text{out}} \).
Setting

Assumptions

\[ Y \sim P_{\text{out}} \left( \cdot \mid \Phi X^* \right) \]

- **Asymptotic regime:** \( n \to \infty, \quad m/n \to \alpha > 0. \)
- \( X^* = (X_1^*, \ldots, X_n^*) \text{i.i.d.} \sim P_0, \quad \mathbb{E}_{P_0} X^2 = \rho. \)
- \( (\Phi_{i,j}) \) are independent, \( \begin{cases} 
\mathbb{E} \Phi_{i,j} = 0 \\
\mathbb{E} \Phi_{i,j}^2 = 1/n \\
\sup_{i,j} \mathbb{E} |\Phi_{i,j}|^3 \text{ remains bounded.}
\end{cases} \)
- \( \mathbb{E} \left[ |Y_\mu|^{2+\epsilon} \right] \) remains bounded, for some \( \epsilon > 0. \)
- \( x \in \mathbb{R} \mapsto P_{\text{out}} (\cdot \mid x) \) is continuous almost everywhere.
- \( P_{\text{out}} \) has to be “regularized” by some (small) Gaussian noise: \( \forall x \in \mathbb{R}, \quad \text{“} P_{\text{out}} (\cdot \mid x) = \tilde{P}_{\text{out}} (\cdot \mid x) + \mathcal{N} (0, \sigma^2) \text{”}) \), where \( \sigma > 0. \)
- If \( P_{\text{out}} \) takes values in \( \mathbb{N} \), no need for regularization (\( \sigma = 0 \)).

The statistician knows the model, i.e. \( P_0 \) and \( P_{\text{out}} \).
Information-theoretic study

The mutual information

Posterior distribution $P(X^*|Y, \Phi)$:

$$P(x|Y, \Phi) = \frac{1}{\mathcal{Z}_n} P_0^{\otimes n}(x) \prod_{\mu=1}^{m} P_{\text{out}}(Y_\mu | \langle \Phi_\mu, x \rangle)$$

where $\mathcal{Z}_n$ is the appropriate normalization.
Information-theoretic study

The mutual information

Posterior distribution $P(\mathbf{X}^*|\mathbf{Y}, \Phi)$:

$$P(\mathbf{x}|\mathbf{Y}, \Phi) = \frac{1}{\mathcal{Z}_n} P_0^{\otimes n}(\mathbf{x}) \prod_{\mu=1}^{m} P_{\text{out}}(Y_\mu | \langle \Phi_\mu, \mathbf{x} \rangle)$$

where $\mathcal{Z}_n$ is the appropriate normalization. The free energy is

$$f_n = -\frac{1}{n} \mathbb{E} \log \mathcal{Z}_n = -\frac{1}{n} \mathbb{E} \left[ \log \int_{\mathbf{x} \in \mathbb{R}^n} dP_0^{\otimes n}(\mathbf{x}) \prod_{\mu=1}^{m} P_{\text{out}}(Y_\mu | \langle \Phi_\mu, \mathbf{x} \rangle) \right]$$
Information-theoretic study

The mutual information

Posterior distribution \( P(X^*|Y, \Phi) \):

\[
P(x|Y, \Phi) = \frac{1}{Z_n} P_0^{\otimes n}(x) \prod_{\mu=1}^{m} P_{\text{out}}(Y_{\mu} | \langle \Phi_{\mu}, x \rangle)
\]

where \( Z_n \) is the appropriate normalization. The free energy is

\[
f_n = -\frac{1}{n} \mathbb{E} \log Z_n = -\frac{1}{n} \mathbb{E} \left[ \log \int_{x \in \mathbb{R}^n} dP_0^{\otimes n}(x) \prod_{\mu=1}^{m} P_{\text{out}}(Y_{\mu} | \langle \Phi_{\mu}, x \rangle) \right]
\]

Equivalently, we are going to study the mutual information:

\[
\frac{1}{n} I(X^*; Y|\Phi) = f_n + \text{Constant} + o_n(1).
\]
Limiting expression for the mutual information

“Replica Symmetric” formula

Theorem

\[
\frac{1}{n} I(X^*; Y | \Phi) \xrightarrow{n \to \infty} \inf_{q \in [0, \rho]} \sup_{r \geq 0} \left\{ I_{P_0}(r) + \alpha I_{P_{\text{out}}}(q) - \frac{r}{2}(\rho - q) \right\}
\]

Example: Linear regression

\[ Y = \Phi X^* + \sigma Z. \]
Limiting expression for the mutual information

“Replica Symmetric” formula

Theorem

\[ \lim_{n \to \infty} \frac{1}{n} I(X^*; Y \mid \Phi) = \inf_{q \in [0, \rho]} \sup_{r \geq 0} \left\{ I_{P_0}(r) + \alpha I_{P_{\text{out}}}(q) - \frac{r}{2}(\rho - q) \right\} \]

Example: Linear regression

- \( Y = \Phi X^* + \sigma Z. \)
- “Tanaka formula”, proved by Barbier et al., 2016 and Reeves and Pfister, 2016.
Limiting expression for the mutual information

“Replica Symmetric” formula

Theorem

\[
\frac{1}{n} I(X^*; Y | \Phi) \xrightarrow{n \to \infty} \inf_{q \in [0, \rho]} \sup_{r \geq 0} \left\{ I_{P_0}(r) + \alpha I_{P_{\text{out}}}(q) - \frac{r}{2}(\rho - q) \right\}
\]

Example: ‘planted’ perceptron.

- \( Y = \text{sign}(\Phi X^*) \), where 
  \( X^*_1, \ldots, X^*_n \) i.i.d. \( \mathcal{U}(+1, -1) \).

- \( S_n = \{ x \mid \forall \mu, \text{sign}(\Phi_\mu x) = Y_\mu \} \)

- \( \frac{1}{n} I(X^*; Y | \Phi) = \log 2 - \frac{1}{n} \mathbb{E}[\log \#S_n] \).

- Formula obtained by Gardner and Derrida, 1989.
Limiting expression for the mutual information

“Replica Symmetric” formula

Theorem

\[
\lim_{n \to \infty} \frac{1}{n} I \left( \mathbf{X}^*; \mathbf{Y} \mid \Phi \right) = \inf_{q \in [0, \rho]} \sup_{r \geq 0} \left\{ I_{P_0}(r) + \alpha I_{P_{\text{out}}}(q) - \frac{r}{2}(\rho - q) \right\}
\]

Example: ‘planted’ perceptron.

\[ Y = \text{sign} (\Phi \mathbf{X}^*), \text{ where} \]
\[ X_1^*, \ldots, X_n^* \sim \text{i.i.d.} \mathcal{U}(+1, -1). \]

\[ S_n = \{ \mathbf{x} \mid \forall \mu, \text{sign}(\Phi \mu \mathbf{x}) = Y_\mu \} \]

\[ \frac{1}{n} I(\mathbf{X}^*; \mathbf{Y} \mid \Phi) = \log 2 - \frac{1}{n} \mathbb{E} \left[ \log \# S_n \right]. \]

Formula obtained by Gardner and Derrida, 1989.
Two scalar inference channels

Explanation of the formula

Recall: \( Y \sim P_{\text{out}}(\cdot | \Phi X^*) \)

\[
\frac{1}{n} I(X^*; Y | \Phi) \xrightarrow{n \to \infty} \inf_{q \in [0, \rho]} \sup_{r \geq 0} \left\{ I_{P_0}(r) + \alpha \mathcal{I}_{P_{\text{out}}}(q) - \frac{r}{2} (\rho - q) \right\}
\]

Additive Gaussian channel

\[
I_{P_0}(r) = I(X^*; \sqrt{r} X^* + Z)
\]

where \( X^* \sim P_0 \) and \( Z \sim \mathcal{N}(0, 1) \).
Two scalar inference channels

Explanation of the formula

Recall: $Y \sim P_{\text{out}}(\cdot \mid \Phi X^*)$

$$\frac{1}{n} I(X^*; Y \mid \Phi) \xrightarrow{n \to \infty} \inf_{q \in [0, \rho]} \sup_{r \geq 0} \left\{ I_{P_0}(r) + \alpha I_{P_{\text{out}}}(q) - \frac{r}{2}(\rho - q) \right\}$$

Additive Gaussian channel

$I_{P_0}(r) = I(X^*; \sqrt{r}X^* + Z)$

where $X^* \sim P_0$ and $Z \sim \mathcal{N}(0, 1)$.

Non-linear Gaussian retrieval

$I_{P_{\text{out}}}(q) = I(W^*; Y(q) \mid V)$

where $V, W^*$ i.i.d. $\mathcal{N}(0, 1)$ and

$Y(q) \sim P_{\text{out}}(\cdot \mid \sqrt{q}V + \sqrt{\rho - q}W^*)$
Proof technique

The interpolation method

In the spirit of Talagrand’s interpolation scheme for the perceptron.

\[ t = 0 \quad \quad 0 < t < 1 \quad \quad t = 1 \]

\[ \mathbf{Y} \sim P_{\text{out}}(\cdot | \Phi \mathbf{X}^*) \]

\[ \begin{cases} \mathbf{Y} \sim P_{\text{out}}(\cdot | \mathbf{S}_t) \\
\mathbf{Y}' = \sqrt{rt}\mathbf{X}^* + \mathbf{Z} \end{cases} \]

\[ \begin{cases} \mathbf{Y} \sim P_{\text{out}}(\cdot | \sqrt{q}\mathbf{V} + \sqrt{\rho-q}\mathbf{W}^*) \\
\mathbf{Y}' = \sqrt{r}\mathbf{X}^* + \mathbf{Z} \end{cases} \]

\[ \mathbf{S}_t = \sqrt{1-t}\Phi \mathbf{X}^* + \sqrt{t}(\sqrt{q}\mathbf{V} + \sqrt{\rho-q}\mathbf{W}^*) \]

\[ \frac{1}{n} I(\mathbf{X}^*; \mathbf{Y} | \Phi) \quad \quad f_n(t) \quad \quad I_{P_0}(r) + \frac{m}{n} I_{P_{\text{out}}}(q) \]
Proof technique
The interpolation method

In the spirit of Talagrand’s interpolation scheme for the perceptron.

**Goal**: show that \( f'_{n}(t) \simeq \frac{r}{2}(\rho - q) \).
Interpolation method

Derivative of the interpolating mutual information

\[ f'_n(t) = \mathbb{E} \left[ \left( \frac{1}{n} \sum_{i=1}^{n} x_i^{(t)} X_i^* - q \right) \text{(Bounded term)} \right] + \frac{r}{2}(\rho - q) \]

where \( x^{(t)} \sim P(X^* | \text{Observations at time } t) \).
Interpolation method

Derivative of the interpolating mutual information

\[ f'_n(t) = \mathbb{E} \left[ \left( \frac{1}{n} \sum_{i=1}^{n} x_i^{(t)} X_i^* - q \right) \left( \text{Bounded term} \right) \right] + \frac{r}{2} (\rho - q) \]

where \( x^{(t)} \sim P(X^* | \text{Observations at time } t) \).

\[ \frac{1}{n} \sum_{i=1}^{n} x_i^{(t)} X_i^* \]

concentrates around some value, and then choose \( q \) to be equal to this value.
Interpolation method

Derivative of the interpolating mutual information

\[ f_n'(t) = \mathbb{E} \left[ \left( \frac{1}{n} \sum_{i=1}^{n} x_i^{(t)} X_i^* - q \right) \text{(Bounded term)} \right] + \frac{r}{2} (\rho - q) \]

where \( x^{(t)} \sim P(X^* | \text{Observations at time } t) \).

- We have to show that the overlap

\[ \frac{1}{n} \sum_{i=1}^{n} x_i^{(t)} X_i^* \]

concentrates around some value, and then choose \( q \) to be equal to this value.

- In the case of Bayes-optimal inference problems this is true under mild assumptions: Montanari, 2008, Korada and Macris, 2010.

- More details about the techniques in Jean Barbier’s talk on Saturday.
Theorem

For almost all $\alpha > 0$, the infimum of the “Mutual Information formula” admits a unique minimizer $q_*(\alpha)$ and

$$\lim_{n \to \infty} \left| \frac{1}{n} \sum_{i=1}^{n} x_i X_i^* \right| = q_*(\alpha), \quad \text{in probability},$$

where $x \sim P(X^* = \cdot | \Phi, Y)$ independently of everything else.

One deduces:

$$\text{MMSE}_n(\alpha) := \frac{1}{n^2} \mathbb{E} \left\| X^* X^*^T - \mathbb{E} [X^* X^*^T | \Phi, Y] \right\|^2 \xrightarrow{n \to \infty} \rho^2 - q_*(\alpha)^2$$
Algorithmic analysis

Generalized Approximate Message Passing (GAMP)

- Generalization of AMP (Donoho et al., 2009) introduced by Rangan, 2011. Iterative algorithm: produces estimates $\hat{x}^0, \ldots, \hat{x}^t$.

\[ \lim_{n \to \infty} \rho - 2 \alpha \sigma^2 P_{\text{out}}(q_t) \]
\[ \text{where } q_t \text{ is given by the recursion } (q_0 = 0): \]
\[
\begin{align*}
q_{t+1} &= \rho - 2 \int P_0(r_t) \, dr_t \\
r_t &= -2 \alpha \int P_{\text{out}}(q_t) \, dq_t
\end{align*}
\]

GAMP converges to a stationary point $(q_{\text{alg}}, r_{\text{alg}})$ of the MI formula and if $q_{\text{alg}} = q^*(\alpha)$, then GAMP achieves the MMSE! 

Main belief: GAMP is optimal among all polynomial-time algorithms.
Algorithmic analysis

Generalized Approximate Message Passing (GAMP)

- Generalization of AMP (Donoho et al., 2009) introduced by Rangan, 2011. Iterative algorithm: produces estimates $\hat{x}^0, \ldots, \hat{x}^t$.
- Its performance can be rigorously tracked:

\[
\frac{1}{n^2} \mathbb{E} \left\| X^* X^{*T} - \hat{x}^t \hat{x}^{tT} \right\|^2 \xrightarrow{n \to \infty} \rho^2 - (q^t)^2
\]

where $q^t$ is given by the recursion ($q^0 = 0$):

\[
\begin{align*}
q^{t+1} &= \rho - 2I'_{P_0}(r^t) \\
r^t &= -2\alpha I'_{P_{out}}(q^t)
\end{align*}
\]

- GAMP converges to a stationary point $(q_{\text{alg}}, r_{\text{alg}})$ of the MI formula and if $q_{\text{alg}} = q_*(\alpha)$, then GAMP achieves the MMSE!
Algorithmic analysis

Generalized Approximate Message Passing (GAMP)

- Generalization of AMP (Donoho et al., 2009) introduced by Rangan, 2011. Iterative algorithm: produces estimates $\hat{x}^0, \ldots, \hat{x}^t$.
- Its performance can be rigorously tracked:

State evolution, Javanmard and Montanari, 2013

$$\frac{1}{n^2} \mathbb{E} \| \mathbf{X}^* \mathbf{X}^* \mathbf{1} - \hat{\mathbf{x}}^t \hat{\mathbf{x}}^t \mathbf{1} \|^2 \xrightarrow{n \to \infty} \rho^2 - (q^t)^2$$

where $q^t$ is given by the recursion ($q^0 = 0$):

$$\begin{cases}
q^{t+1} = \rho - 2I_P'(r^t) \\
r^t = -2\alpha I_{P_{out}}'(q^t)
\end{cases}$$

- GAMP converges to a stationary point $(q^{\text{alg}}, r^{\text{alg}})$ of the MI formula and if $q^{\text{alg}} = q_*(\alpha)$, then GAMP achieves the MMSE!
- **Main belief**: GAMP is optimal among all polynomial-time algorithms.
Phase diagrams: warm-up

Linear model

\[ Y = \Phi X^*, \quad P_0 = \rho N(0, 1) + (1 - \rho)\delta_0 \]

Phase diagram from Krzakala et al., 2012
ReLu channel

\[ \mathbf{Y} = \text{ReLu}(\Phi \mathbf{X}^*), \quad \text{ReLu}(x) = x 1(x \geq 0), \quad P_0 = \rho \mathcal{N}(0, 1) + (1 - \rho)\delta_0 \]
Absolute value channel

\[ Y = |\Phi X^*|, \quad P_0 = \rho N(0, 1) + (1 - \rho)\delta_0 \]
Symmetric door channel

\[ Y = 1(\Phi X^* \in [-K, K]), \quad P_0 = \frac{1}{2} \delta_{-1} + \frac{1}{2} \delta_{+1} \]
A learning problem

A different point of view

- The points \( \{(\Phi_1, Y_1), \ldots, (\Phi_m, Y_m)\} \) can be seen as data generated by some relation \( Y \sim P_{\text{out}}(\cdot|\Phi X^*) \).

- **Question**: How difficult is it to learn this relation?
A learning problem

A different point of view

- The points \(\{(\Phi_1, Y_1), \ldots, (\Phi_m, Y_m)\}\) can be seen as data generated by some relation \(Y \sim P_{out}(\cdot | \Phi X^*)\).

- **Question**: How difficult is it to learn this relation?

- What is the optimal generalization error

\[
\mathcal{E}_{n}^{\text{gen}} = \min_{\hat{\theta}} \mathbb{E} \left[ (Y^{(\text{new})} - \hat{\theta}(\Phi^{(\text{new})}; Y, \Phi))^2 \right]
\]

where \(Y^{(\text{new})} \sim P_{out}(\cdot | \langle \Phi^{(\text{new})}, X^* \rangle)\) is a new sample.
A learning problem

A different point of view

- The points \( \{(\Phi_1, Y_1), \ldots, (\Phi_m, Y_m)\} \) can be seen as data generated by some relation \( Y \sim P_{out}(\cdot | \Phi X^*) \).

- **Question**: How difficult is it to learn this relation?

- What is the optimal generalization error

\[
E^\text{gen}_n = \min_{\hat{\theta}} \mathbb{E} \left[ (Y^{(\text{new})} - \hat{\theta}(\Phi^{(\text{new})}; Y, \Phi))^2 \right]
\]

where \( Y^{(\text{new})} \sim P_{out}(\cdot | \langle \Phi^{(\text{new})}, X^* \rangle) \) is a new sample.

Theorem

\[
E^\text{gen}_n \xrightarrow{n \to \infty} E(q^*(\alpha))
\]

where

\[
E(q) = \mathbb{E} \left[ (Y(q) - \mathbb{E}[Y(q)|V])^2 \right]
\]

Recall the second scalar channel: \( Y^{(q)} \sim P_{out}(\cdot | \sqrt{q}V + \sqrt{\rho - q}W) \), \( V, W \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1) \).
Classification: the perceptron

\[ Y = \text{sign} (\Phi X^*) , \quad P_0 = \frac{1}{2} \delta_{-1} + \frac{1}{2} \delta_{+1} \]

Computed by Györgyi, 1990 and also Seung et al., 1992:
Regression: phase retrieval

\[ Y = |\Phi X^*|, \quad P_0 = \frac{1}{2} \delta_{-1} + \frac{1}{2} \delta_{+1} \]
Classification: the symmetric door

\[ Y = 1(\Phi X^* \in [-K, K]), \quad P_0 = \frac{1}{2} \delta_{-1} + \frac{1}{2} \delta_{+1} \]

\begin{center}
\begin{tikzpicture}
\draw[->] (0,0) -- (0,2.5); \draw[->] (0,0) -- (2.5,0);
\node at (2.5,0.5) {\textbf{SE}};
\node at (2.5,0.25) {\textbf{Bayes-optimal}};
\node at (2.5,0.01) {GAMP, n=10^4};
\node at (0.1,0.2) {\textbf{NeuralNet}};
\end{tikzpicture}
\end{center}
Thank you for your attention.

Any questions?
References I

- Javanmard, Adel and Andrea Montanari (2013). “State evolution for general approximate message passing algorithms, with applications to spatial coupling”. In: Information and Inference, iat004.
References II