Bayesian Field Theory

Jörg Lemm

Institut für Theoretische Physik I
Universität Münster

lemm@uni-muenster.de
http://pauli.uni-muenster.de/~lemm

Training data: e.g., $h(x_1) = 5$

Test data: $h(x_2) = ?$

Dependencies (prior information, rules):
e.g., "smoothness": $h(x_1) - h(x_2) = 2$

Generalization: $h(x_2) = 7$

$\Rightarrow$

Two aspects of generalization:

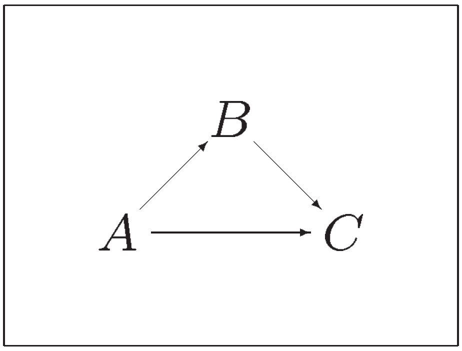
☞ algorithmic aspect: reformulation of knowledge to find answers for test set given training data and dependencies.

☞ control aspect: control of dependencies between test and training data (prior information = posterior control)

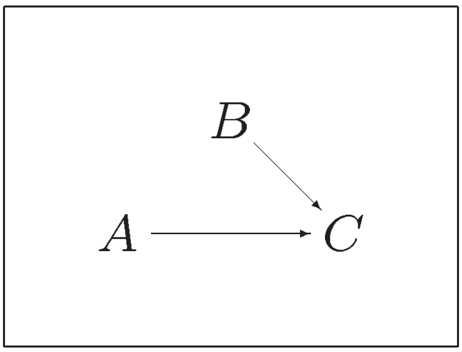
Joint and conditional probabilities

$$p(A,B) = p(B)p(A|B) = p(A)p(B|A)$$

$$p(A,B,C) = p(A)p(B|A)p(C|A,B)$$



$$p(A,B,C) = p(A)p(B)p(C|A,B)$$



$$p(A,B) = p(A|B)p(B) = p(B|A)p(A) \Rightarrow$$

Bayes' rule

$$p(B|A) = \frac{p(A|B)p(B)}{p(A)}$$

$$p(\text{model}|\text{data}) = \frac{p(\text{data}|\text{model}) * p(\text{model})}{p(\text{data})}$$

$$\boxed{\text{posterior} = \frac{\text{likelihood } * \text{ prior}}{\text{evidence}}}$$

1. **hidden variables** (not directly measurable)

   $$h \in \mathcal{H}$$

   (*model states*, *possible state of Nature*)

2. **visible variables**

   a. **independent variables** (describing kinds of measurement)

   $$x \in \mathcal{X}$$

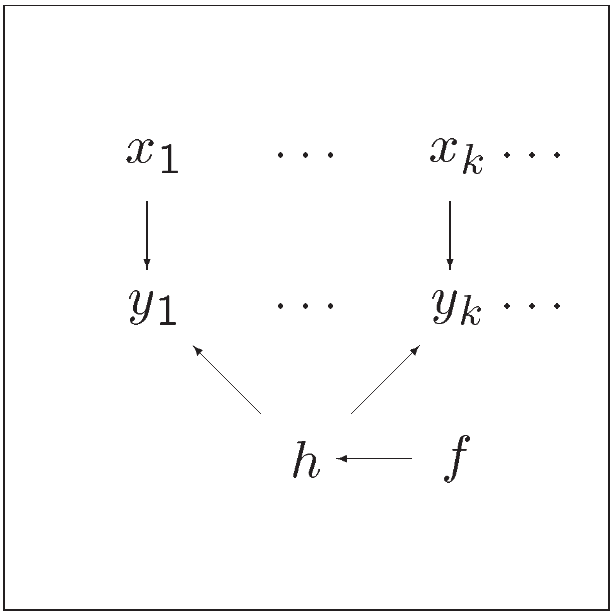
   b. **dependent variables** (responses)

   $$y \in \mathcal{Y}$$

   c. **determining variables** (state of knowledge)

   $$f \in \mathcal{F}$$

   (determining probability $p(h|f)$)



Graphical representation of a probabilistic model factorizing according to

$$p(x, y, h|f) = p(h|f) \prod_k p(x_k)p(y_k|x_k, h).$$

1. **relevant** or **test data** $D_R = (x_R, y_R)$, $x_R \in \mathcal{X}_R$, $y_R \in \mathcal{Y}_R$

2. **available data** D contribute to our *state of knowledge* about Nature.

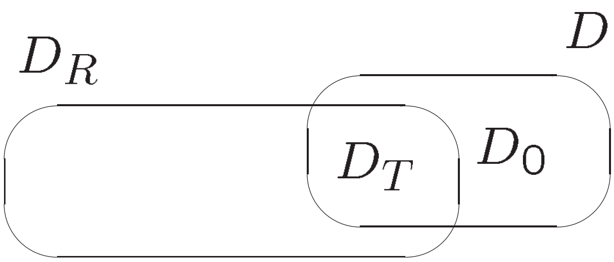
   a. **training data** $D_T = \{D_{T,i} | 1 \le i \le n\} = \{(x_i, y_i) | 1 \le i \le n\} = (x_T, y_T)$, empirical sample of D_R,

   b. **prior data** $D_0 = \{(x_0, y_0), f, S\}$

      i. **measured prior**, (x_0, y_0)

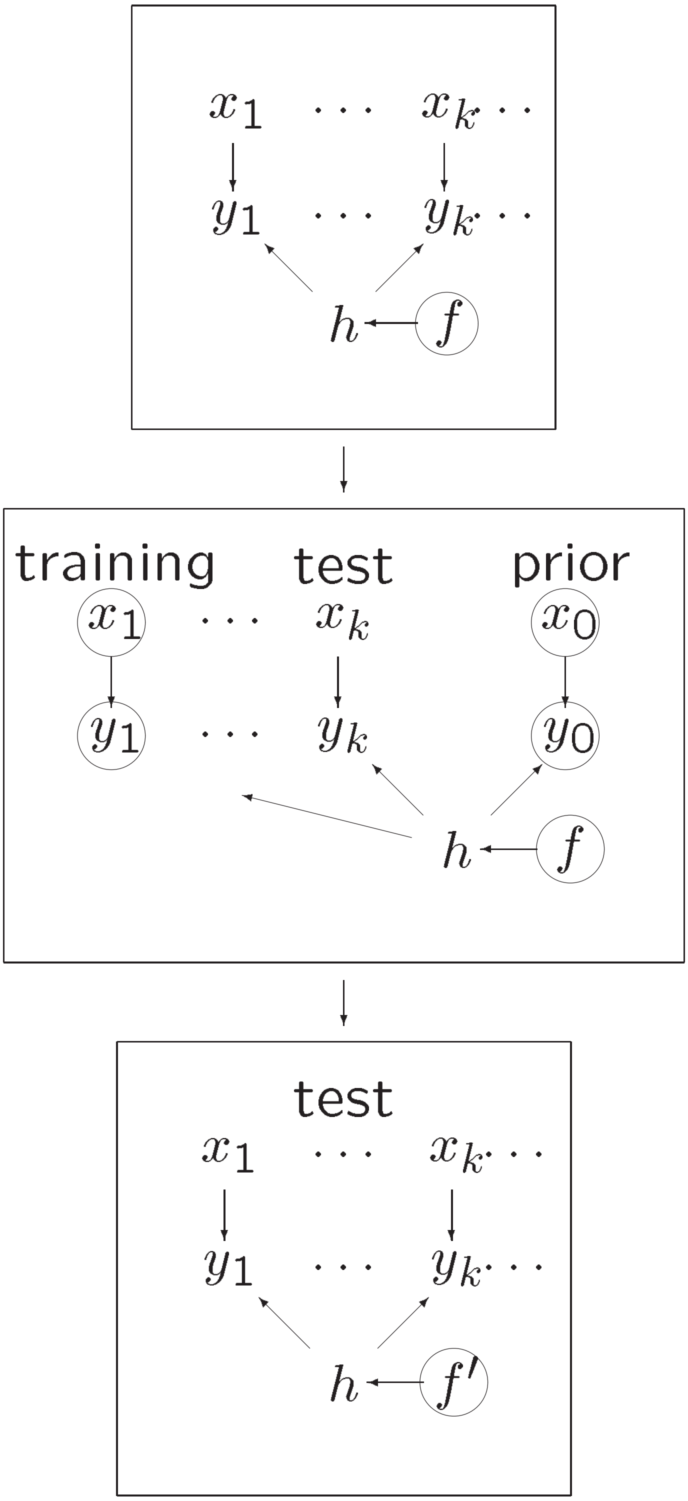
      ii. **generative prior** f, knowledge/control concerning the preparation of f

      iii. **structural prior** S knowledge/control concerning the specified dependency structure of the model



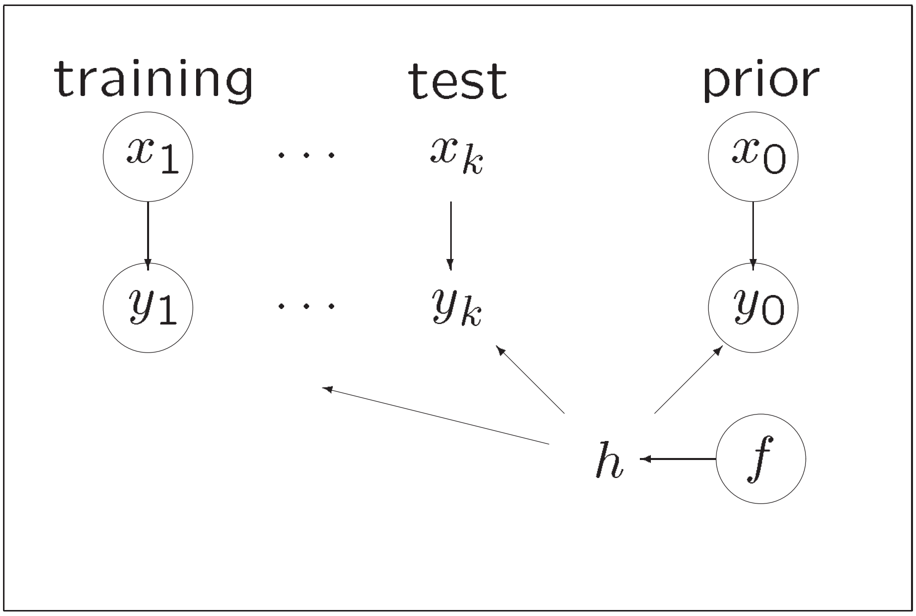
Typical relation between relevant data D_R, training data $D_T \subset D_R$, prior data D_0 and available data $D = D_T \cup D_0$.

Learning



$x_1 \cdots x_k \cdots$
$y_1 \cdots y_k \cdots$
h
f
training
test
prior
$x_1 \cdots x_k$
x_0
$y_1 \cdots y_k$
y_0
h
f
test
$x_1 \cdots x_k \cdots$
$y_1 \cdots y_k \cdots$
h
f'

From data to predictive density



Model with measured prior:

$$p(x, y, h|f) =$$

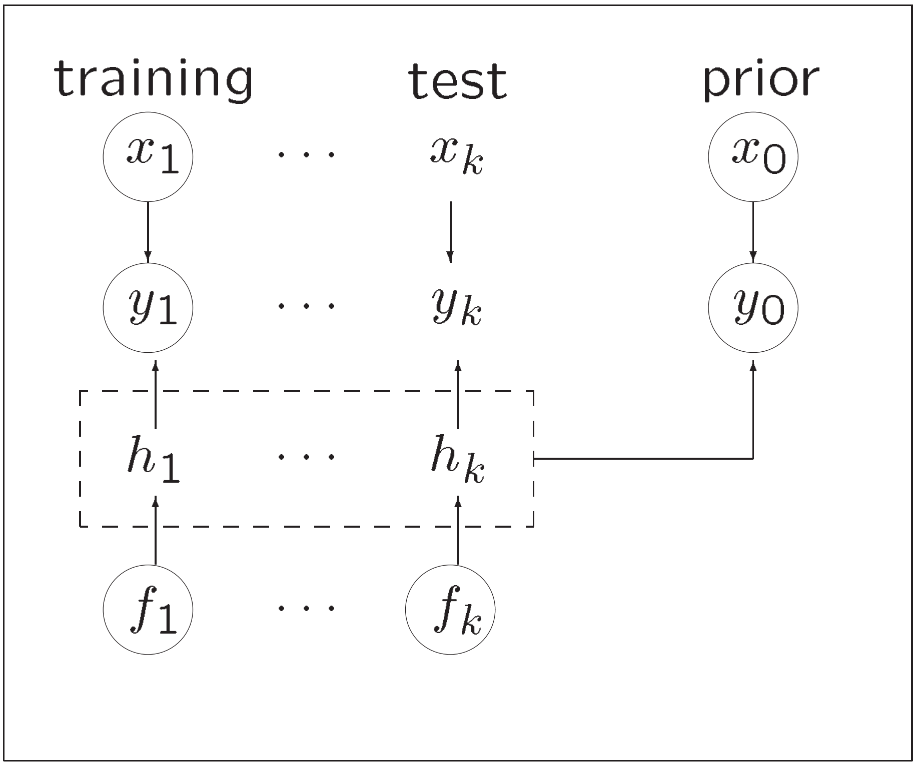
$$p(h|f)p(x_0)p(y_0|x_0, h) \prod_{k=1} p(x_k)p(y_k|x_k, h).$$

Predictive density for test (relevant) data:

$$\boxed{p(y_R|x_R, f'(D, f)\,) = \int dh\; p(y_R|x_R, h)\, p(h|D)}$$

requires calculation of **posterior density**:

$$\boxed{p(h|D) \propto p(y_D|x_D, h)\, p(h|f)}$$



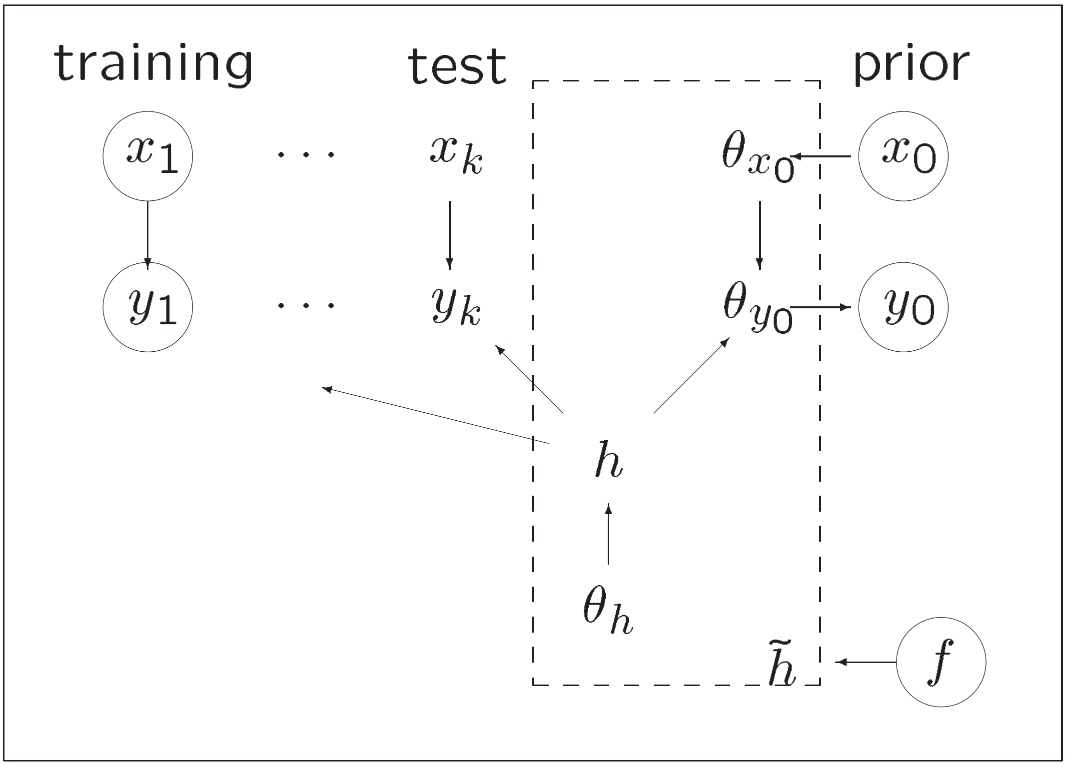
Graphical representation of a factorial prior

$$p(x, y, h|f) =$$

$$p(x_0)p(y_0|x_0, h) \prod_{k=1} p(h_k|f_k)p(x_k)p(y_k|x_k, h_k).$$

Without y_0 depending on h_k no generalization is possible from training data $D_{l \neq k}$ to test data D_k.

Input, output, and generative noise



$$p(x, y, \tilde{h}) = p(x, y, h, \theta_{x_0}, \theta_{y_0}, \theta_h)$$

$$= p(\theta_h)p(h|\theta_h)p(x_0)p(y_0|\theta_{x_0})p(\theta_{y_0}|\theta_{x_0}, h)$$

$$\times p(\theta_{x_0}|x_0) \prod_{k=1} p(x_k)p(y_k|x_k, h).$$

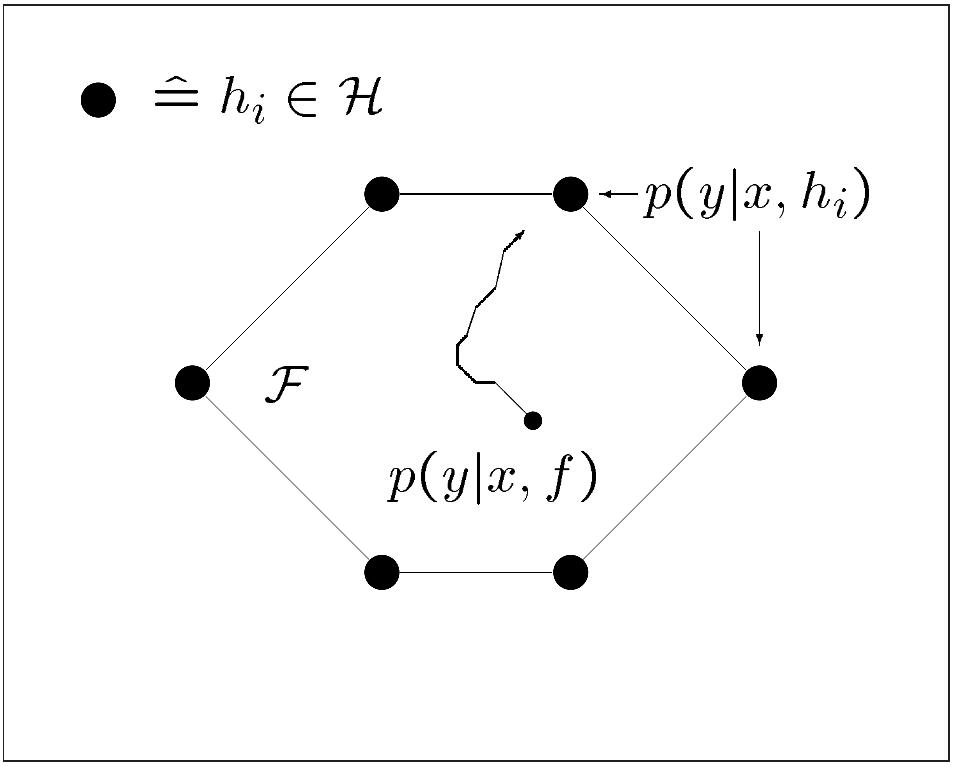
$$p(h|D) \propto \int d\theta_{x_0}\, d\theta_{y_0}\, d\theta_h p(y_T|x_T, h)p(y_0|\theta_{y_0})$$

$$\times p(\theta_{y_0}|\theta_{x_0}, h)p(\theta_{x_0}|x_0)p(h|\theta_{f_0})p(\theta_{f_0}).$$

Learning: Schematic model

Data dependent state of knowledge $f = f(D)$, learning means updating posterior $p(h_i|f)$ to obtain new predictive density

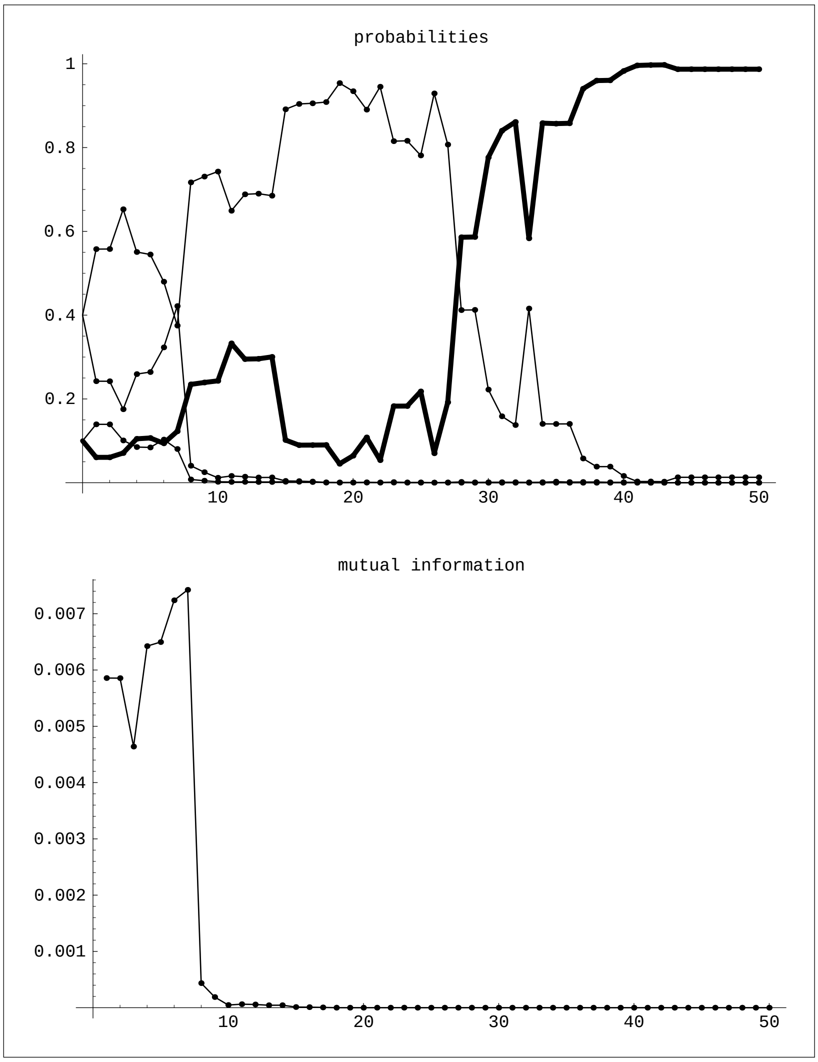
$$p(y|x,f) = \sum_i p(y|x,h_i)\,p(h_i|f)$$



Mutual Information I

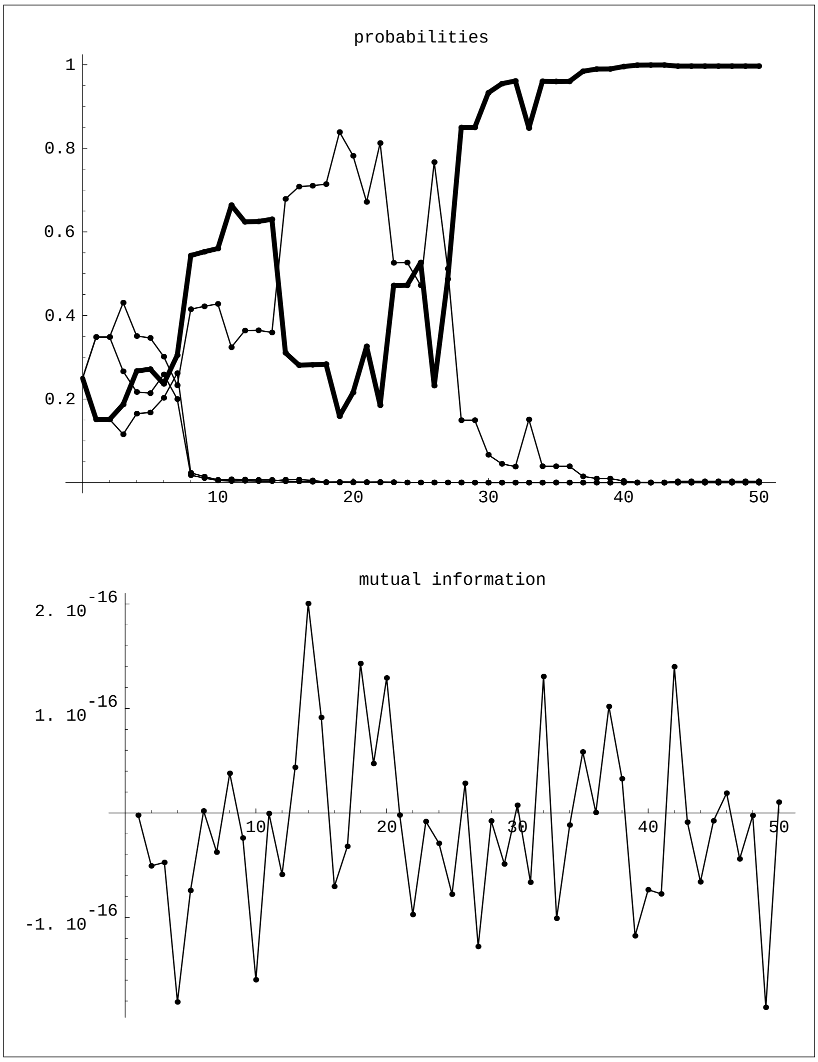
$$M(y) = \int \left(\prod_i dy_i \right) p(y|x,f) \ln \frac{p(y|x,f)}{\prod_j p(y_j|x_j,f)}$$

$$M(y) = \sum_i H(y_i) - H(y)$$



Factorial prior, pure local learning

$$M(y) = \sum_i H(y_i) - H(y) = 0$$



Assume we have to choose in situation x between possible **actions** $a(x) \in \mathcal{A}$ and are able to define a

loss function (loss suffered if action a is chosen in situation x and outcome y occurs)

$$l(x, y, a(x))$$

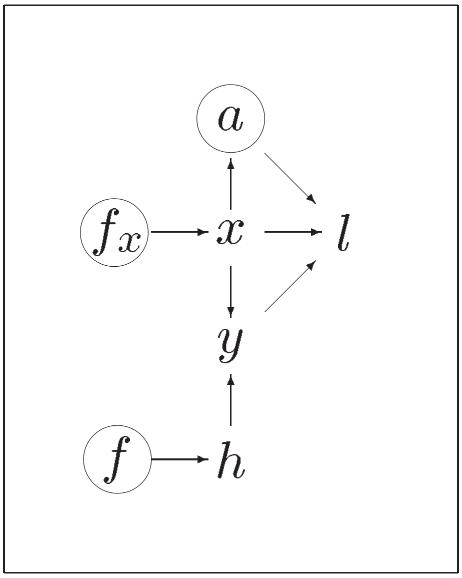
E.g. for density estimation: $-\ln p(y|x, a)$ for regression: $(y - a(x))^2$, $|y - a(x)|$, $1 - \delta_{y,a(x)}$

Aim: Minimize functional of $p(l|a, f)$, e.g.
expected risk

$$r(a, f) = \int dl \; l \; p(l|a, f)$$

$$= \int dh \, dx \, dy \, l(x, y, a(x)) \, p(x) \, p(y|x, h) \, p(h|f)$$

Decision model



deterministic loss

$$p(l|x,y,a) = \delta(l - l(x,y,a))$$

deterministic action

$$p(a|x) = \delta(a - a(x))$$

loss posterior under action a

$p(l|a,f) =$

$\int dh\, dx\, dy\, p(l|x,y,a(x))\, p(x|f_x)\, p(y|x,h)\, p(h|f)$

expected risk (= expected loss)

$r(a,h) =$

$\int dh\, dx\, dy\, l(x,y,a(x))\, p(x|f_x)\, p(y|x,h)\, p(h|f)$

Approximation methods to calculate the expected risk $r(a, f)$ include for example:

- Sampling (Monte Carlo)
- Saddle point approximation (perturbation theory)

Saddle point approximation
= maximum a posteriori approximation (MAP)

$$h^* = \text{argmax}_h \, p(h|f) = \text{argmin}_h(-\ln p(h|f))$$

defining the **error functional** to be minimized

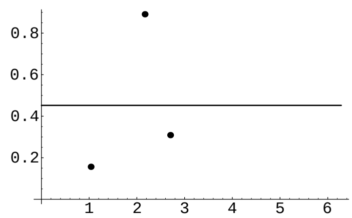
$$E(h) = -\ln p(h|f))$$

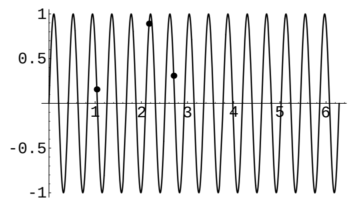
predictive density becomes

$$p(y|x, f)_{\text{MAP}} = p(y|x, h^*)$$

Parameterized h, e.g. mean square error (Non–parametric methods = many parameters, e.g. neural nets)

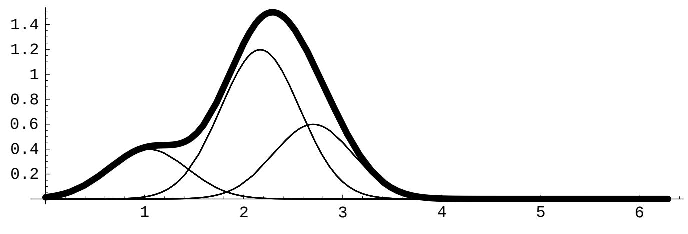
$$E(\xi) = \frac{1}{2}\sum_i (y_i - h(x_i, \xi))^2$$





Regularized error functionals, e.g.

$$E(h) = \frac{1}{2}\sum_i (y_i - h(x_i))^2 + \frac{\lambda}{2}\int dx \left(\frac{dh(x)}{dx}\right)^2$$



Choose possible actions a equal to possible states of knowledge f, or, for MAP, equal to h, i.e., $\mathcal{A} = \mathcal{F}$ (or $\mathcal{A} = \mathcal{H}$) and use **log–loss**,

$$l(x, y, a) = -\ln p(y|x, a).$$

The expected risk becomes up to a constant equal to the (x–averaged) **Kullback-Leibler entropy** (relative entropy, cross–entropy)

$$r(a, f) = -\int dx\, p(x) \int dy\, p(y|x, f) \ln \frac{p(y|x, a)}{p(y|x, f)}$$

It can be shown using Jensen's inequality that the Kullback-Leibler entropy is minimized for

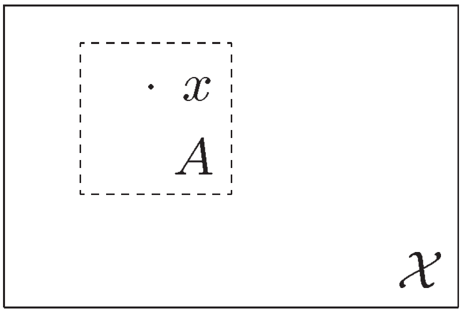
$$p(y|x, a) = p(y|x, f)$$

or for MAP $p(y|x, a) = p(y|x, h^*)$. It follows

$$\begin{aligned} \mathcal{A} &= \mathcal{F} \Rightarrow a^* = f \\ \mathcal{A} &= \mathcal{H} \Rightarrow a^* = h^* \quad \text{for MAP} \end{aligned}$$

Thus, the optimal action (approximation) a^* is already given by the predictive density.

Reparameterizing probability



$x \in A \subseteq \mathcal{X}$,

$p(A) = \int_A dx\, p(x)$,

$p(\mathcal{X}) = 1$.

probability

$$p(x) = \frac{Z(x)}{Z(\mathcal{X})} = e^{-b(x)} = e^{-\beta(F(x)-F(\mathcal{X}))}$$

$$p(A) = \frac{Z(A)}{Z(\mathcal{X})} = e^{-b(A)} = e^{-\beta(F(A)-F(\mathcal{X}))}$$

partition sum (unnormalized probability)

$$Z(A) = \int_A dx\, Z(x)$$

bit number (information)

$$b(A) = -\ln p(A)$$

free energy

$$F(A) = -\frac{1}{\beta} \ln Z(A)$$

conditional average

$$\langle f(x) \rangle_A = \int_A dx \, \frac{p(x)}{p(A)} f(x).$$

(average) information, $I(x) = b(x)$,

$$b(A) = -\langle \ln p(x) \rangle_A$$

energy, $E(x) = F(x)$,

$$E(A) = -\frac{1}{\beta} \langle \ln Z(x) \rangle_A .$$

entropy (conditional information):

$$H(A) = -\langle \ln p(x|A) \rangle_A$$

$$= -\left\langle \ln \frac{p(x)}{p(A)} \right\rangle_A = -\left\langle \ln \frac{Z(x)}{Z(A)} \right\rangle_A$$

$$= -\langle \ln p(x) \rangle_A + \langle \ln p(A) \rangle_A = I(A) - b(A)$$

$$= -\langle \ln Z(x) \rangle_A + \langle \ln Z(A) \rangle_A = \beta E(A) - \beta F(A).$$

Summarizing table

Transformed probability	Averages of canonical random variable	Difference
bit number $b(A) = -\ln p(A)$	information $I(A) = -\langle \ln p(x) \rangle_A$	entropy $H(A) = -\langle \ln p(x\|A) \rangle_A$ $= I(A) - b(A)$
free energy $F(A) = -\dfrac{1}{\beta} \ln Z(A)$	energy $E(A) = -\dfrac{1}{\beta} \langle \ln Z(x) \rangle_A$	entropy $H(A) = -\langle \ln p(x\|A) \rangle_A$ $= \beta(E(A) - F(A))$

Gaussian training data (regression)

$$p(y|x,h) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-h(x))^2}{2\sigma^2}}$$

state h specified by its **regression function**

$$h(x) = \int dy\, y\, p(y|x,h).$$

Gaussian prior data (classical regularization)

$$p(y_0|x_0,h) = (2\pi)^{-\frac{d}{2}} (\det K)^{\frac{1}{2}} e^{-\frac{1}{2}\left\langle h-y_0 \middle| K(x_0) \middle| h-y_0 \right\rangle}$$

Error functional $E(h)$ becomes

$$\frac{1}{2}\sum_i (y_i - h(x_i))^2 + \frac{1}{2}\left\langle h - y_0 \middle| K(x_0) \middle| h - y_0 \right\rangle$$

Choosing, for example

$$K = -\Delta$$

equal to the negative d–dimensional Laplacian (**kinetic energy**) implements a smoothness constraint. Data y_i correspond to an δ–like external field and mass term.

Density operator: $\rho = \sum_i p_i \, |\varphi_i >< \varphi_i|$

$$p(y|x,h) = p(y|\mathbf{x},\rho) = \mathrm{Tr}(\mathbf{P}_{\mathbf{x},y}\rho) = \sum_j^{n_y} < y_j| \, \rho \, |y_j>$$

with $\mathbf{x}|y_j >= y|y_j >, \quad \left\langle y_j|y_i'\right\rangle = \delta_{ji}\delta(y-y')$,

$$\mathbf{P}_{\mathbf{x},y} = \sum_j^{n_y} |y_j >< y_j|$$

Quantum statistics:

$$\rho(\mathbf{H}(\xi)) = \frac{e^{-\beta \mathbf{H}(\xi)}}{\mathrm{Tr}\, e^{-\beta \mathbf{H}(\xi)}}$$

Scattering:

$$p(y|x,h) = \lim_{\substack{t\to\infty \\ t_0\to-\infty}} p(y|\mathbf{x},\rho(t,z(t_0))) =$$

$$\lim_{\substack{t\to\infty \\ t_0\to-\infty}} |\,\langle y(t)|\mathbf{U}(t,t_0)z(t_0)\rangle\,|^2 = |\langle y^0(0)|\mathbf{S}z^0(0)\rangle|^2$$

which defines the scattering operator $\mathbf{S} = \lim_{\substack{t\to\infty \\ t_0\to-\infty}} \mathbf{U}^\dagger_{out}(t,0)\mathbf{U}(t,t_0)\mathbf{U}_{in}(t_0,0)$.

Gaussian processes correspond to quadratic error functionals (classical regularization). We now use (convex) quadratic building blocks to construct general non–convex error functionals going beyond classical regularization

Quadratic concept

$$\boxed{d^2[h] = \langle h - t \,|K|\, h - t\rangle}$$

$t = t(x)$ template (mean, prototype, data)
$K = P_t K_t P_t$, P_t = projector,
$K_t = W^T W = (I - S)^T (I - S)$, symmetric, positive definite.

Example: mean square data term

$$\langle h - t \,|K|\, h - t\rangle = (y_i - h(x_i))^2$$

with $t(x) = y_i$ and $K(x, x') = \delta(x - x')\delta(x - x_i)$.

Example: Smoothness

$$\langle h - t \,|K|\, h - t\rangle = \langle h \,|K|\, h\rangle$$

with $t(x) \equiv 0$ (zero template), and
$K = -\Delta$ (kinetic energy)
$K = \sum_{m=0}^{\infty} (-1)^m \frac{\sigma^{2m}}{m! 2^m} \Delta^m$ (Gaussian–RBF)

Approximating

$$\boxed{t_1 \text{ AND } t_2 \text{ AND } \cdots \text{ AND } t_N}$$

corresponds to an additive error functional

$$\boxed{E = \sum_i^N d_i^2[h]/2 = d^2[h]/2 + E_0}$$

with square distance

$$d^2[h] = \langle h - t \,|K|\, h - t\rangle$$

from template average

$$t = K^{-1}\tilde{t} \text{ with } \tilde{t} = \sum_i^N K_i t_i \text{ and } K = \sum_i K_i$$

and h–independent minimal energy

$$E_0 = \frac{1}{2}\Big(\sum_i^N \langle t_i \,|K_i|\, t_i\rangle - \langle t \,|K|\, t\rangle\Big).$$

The stationarity equation is linear and reads

$$\boxed{0 = \tilde{t} - Kh \Rightarrow h = t.}$$

$$\boxed{E[h] = \left\langle h - t_D \,\middle|\, K_D \,\middle|\, h - t_D \right\rangle + \left\langle h - t_0 \,\middle|\, K_0 \,\middle|\, h - t_0 \right\rangle}$$

$$\left\langle h - t_D \,\middle|\, K_D \,\middle|\, h - t_D \right\rangle = \sum_i^n (y_i - h(x_i))^2 + \text{c}$$

The stationarity equation

$$0 = K_D(h - t_D) + K_0(h - t_0)$$

yields

$$\boxed{h = K_0^{-1} a + t_0,}$$

with $a = K_D(t_D - h)$.

For known covariance $\Sigma = K_0^{-1}$ it is sufficient to solve

$$\boxed{a = [I + K_D K_0^{-1}]^{-1} [K_D(t_D - t_0)]}$$

in the n–dimensional space of functions defined on data points x_i.

Ambiguous data or ambiguous prior

$$\boxed{\text{(data AND } t_1\text{) OR (data AND } t_2\text{)}}$$

$t_1(x)$ and $t_2(x)$ represent two (alternative) function templates (prototypes).

I. Finite temperature/mixture regularization

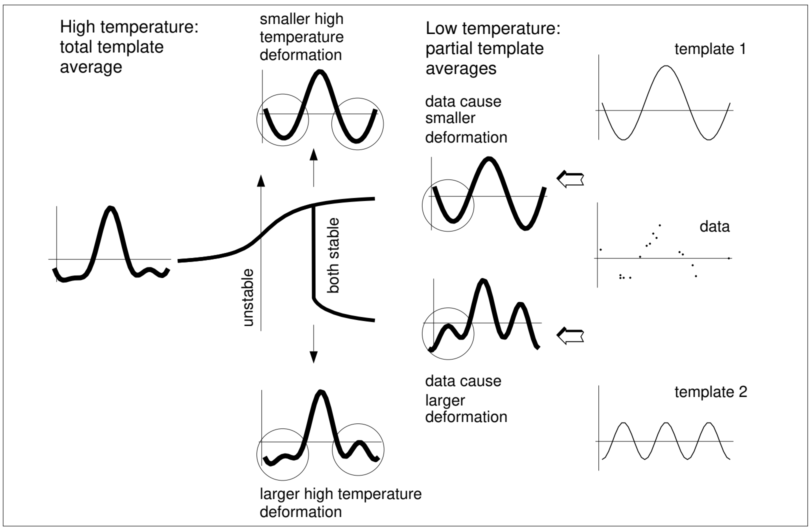
$$\begin{aligned} p^M &= p_1^M + p_2^M \propto e^{-\beta E} \\ &= \boxed{e^{-\beta E_1} + e^{-\beta E_2}} \\ &= e^{-\frac{\beta}{2}\left(d_D^2 + d_1^2\right)} + e^{-\frac{\beta}{2}\left(d_D^2 + d_2^2\right)} \end{aligned}$$

$d_D^2 = \sum_i (y_i - h(x_i))^2$
$d_i^2 = -\left\langle h - t_i \middle| \Delta \middle| h - t_i \right\rangle$(*non–zero* t_i)
The stationarity equation $h = p_1^M t_1 + p_2^M t_2$ becomes (ferromagnet)

$$\boxed{h = t + \tanh\left[\frac{\beta}{2}(E_2 - E_1)\right] \frac{t_1 - t_2}{2}}$$

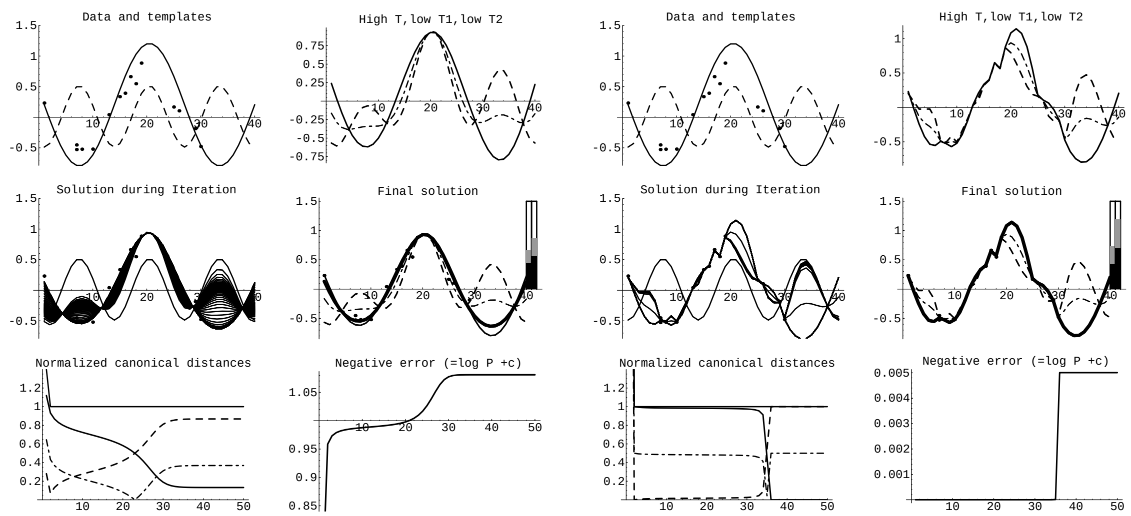
OR: Two template example (II)

Temperature dependence:



Linear problems:

- High temperature limit
- (two) low temperature limits,
- classical regularization ($t_1 = t_2 \equiv 0$).



OR: Two template example (III)

$$\boxed{(\text{data AND } t_1) \text{ OR } (\text{data AND } t_2)}$$

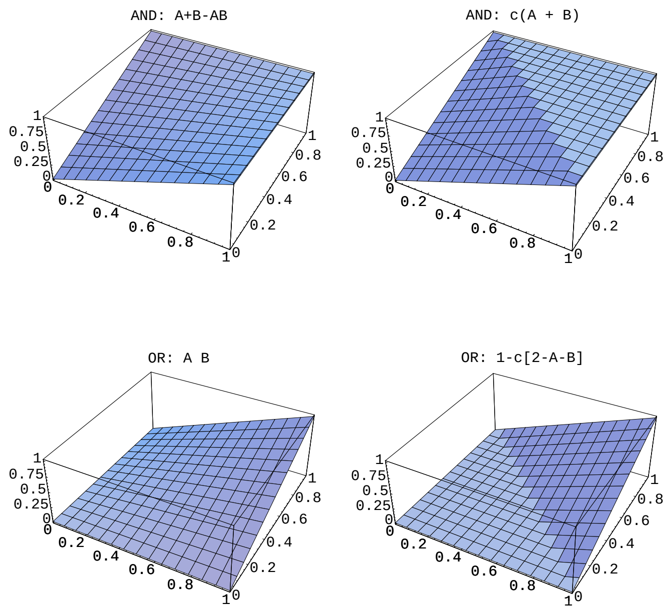
II. Landau–Ginzburg regularization

$$\boxed{E = g(d_D^2 + \beta\, d_1^2\, d_2^2)}$$

(g strictly monotone)
with stationarity equation

$$\boxed{\left[K_D - \beta(d_1^2 + d_2^2)\Delta\right] h = K_D D - \beta\Delta(d_1^2\, t_1 + d_2^2\, t_2)}$$

$$K_D(x, x') = \delta(x - x') \textstyle\sum_i \delta(x - x_i)$$
$$D(x) = \textstyle\sum_n \delta(x - x_i) y_i / (\sum_n \delta(x - x_i))$$



Mixture model:

$$p(h) = \sum_i^{N_i} p(h,i) = \sum_i^{N_i} p_i \, p(h|i),$$
$$p(h|i) = e^{-\beta E_i}/Z_i,$$
$$Z_i = \int dh \, e^{-\beta E_i},$$
$$p_i = e^{-b e_i}/z, \; z = \sum_i e^{-b e_i}$$

yields a spin glas like error functional

$$\boxed{E^M[h] = -\ln\left(\sum_i^{N_i} e^{-\beta E_i[h] - b e_i - \ln Z_i(\beta)}\right),}$$

$$\text{with } E_i = \frac{1}{2} \sum_{j_i}^{n_{j_i}} d_{i j_i}^2.$$

The (nonlinear) stationarity condition becomes

$$\boxed{0 = t^M[h] - K^M[h] h,}$$

with

$$K^M = \sum_i p(h,i) K_i,$$
$$t^M = \sum_i p(h,i) K_i \bar{t}_i = \sum_i p(h,i) \tilde{t}_i.$$

A product implementation of OR ('fuzzy OR') yields models with **polynomial interactions**

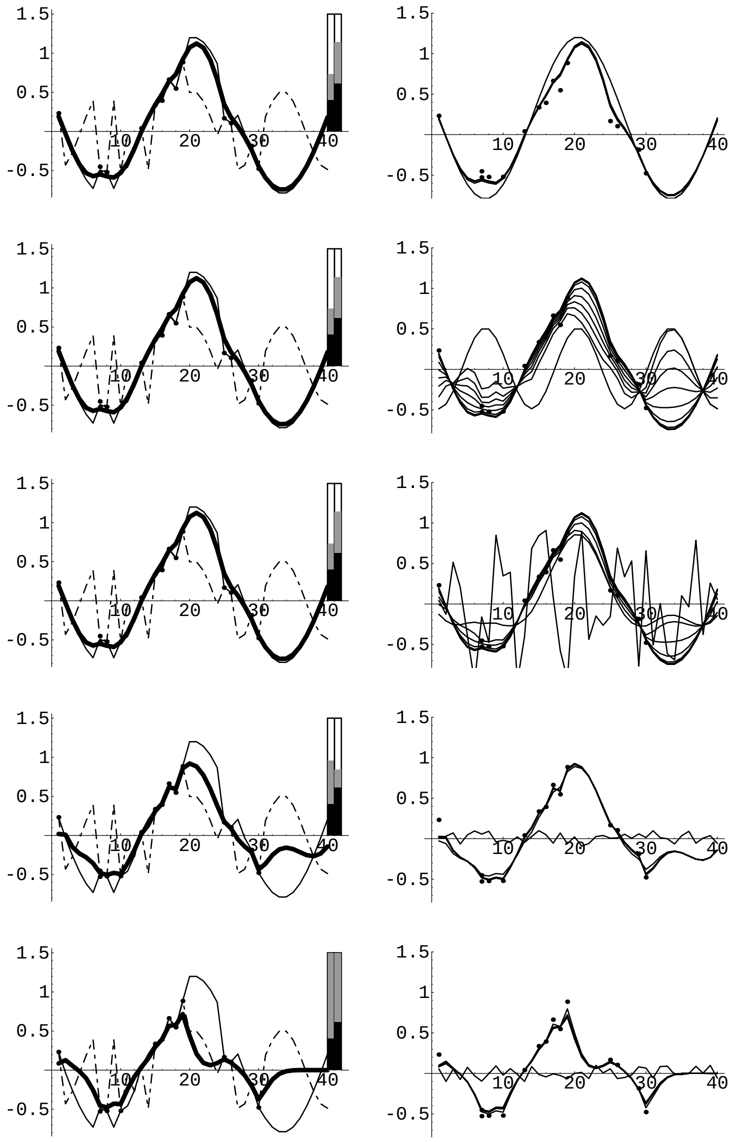
$$\boxed{E^{LG} = \frac{1}{2}\sum_{i=1}^{n_i} \beta_{LG}^{n_{j_i}-1} \prod_{j_i=1}^{n_{j_i}} d_{ij_i}^2,}$$

and (nonlinear) stationarity equation

$$\boxed{0 = t^{LG}[h] - K^{LG}[h]h.}$$

$$\begin{aligned} K^{LG}[h] &= \sum_i \sum_{j_i} M_{ij_i}[h] K_{ij_i}, \\ t^{LG}[h] &= \sum_i \sum_{j_i} M_{ij_i}[h] K_{ij_i} t_{ij_i}, \\ M_{ij_i}[h] &= \beta_{LG}^{n_{j_i}-1} \prod_{k_i \neq j_i} d_{ik_i}^2[h]. \end{aligned}$$

Landau–Ginzburg model: Numerical example



$$\boxed{h^{k+1} = h^k + \eta A^{-1}(t^k[h^k] - K^k[h^k]h^k)}$$

Learning algorithm	Learning matrix
Gradient	$\mathbf{A} = \mathbf{I}$ (Identity)
Jacobi	$\mathbf{A} = \mathbf{D}$ (diagonal)
Gauss–Seidel	$\mathbf{A} = \mathbf{D} + \mathbf{L}$ (low.triang.)
Newton	$\mathbf{A} = -\mathbf{H}$ (neg. Hessian)
Prior–relaxation	$\mathbf{A} = \mathbf{K}$ (EM for mixture)
Massive relaxation	$\mathbf{A} = \mathbf{A}_0 + m^2\mathbf{I}$
Linear boundary	$\mathbf{A} = \mathbf{A}_{II} + \mathbf{A}_{IB} + \mathbf{A}_{BB}$
Gaussian	$\mathbf{A} = \sum_{k=0}^{\infty} \frac{1}{k!}\left(\frac{\mathbf{M}^2}{2\tilde{\sigma}^2}\right)^k = e^{\frac{\mathbf{M}^2}{2\tilde{\sigma}^2}}$

Hessian, mixture model

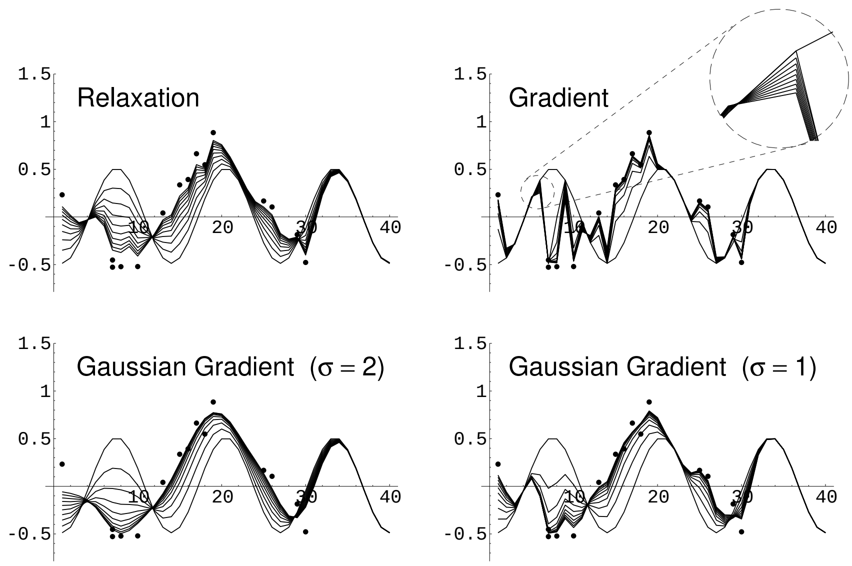
$$H^M = -\beta \sum_i p(h,i) K_i$$
$$+\beta^2 \sum_i p(h,i) |K_i h - t_i\rangle\langle K_i h - t_i|$$
$$-\beta^2 \sum_{ij} p(h,i) |K_i h - t_i\rangle\langle K_j h - t_j| p(h,j).$$

Hessian, Landau–Ginzburg

$$H^L = \sum_i \sum_{j_i} M_{ij_i} K_{ij_i}$$
$$+\sum_i \sum_{j_i} \sum_{k_i} M_{ij_ik_i} |K_{ij_i} h - t_{ij_i}\rangle\langle K_{ij_i} h - t_{ij_i}|$$

with $M_{ij_ik_i} = \prod_{j_i \neq l_i \neq k_i} \beta_F^{n_{j_i}-1} d^2_{il_i}[h]$

Naive gradient vs. gaussian relaxation



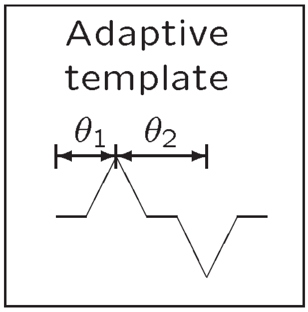
OR–like combinations describe variations of templates. Often those are continuously parameterized (translations, scaling, $\cdots$).

$$\boxed{p^M \propto \int d\theta\, p(\theta) \sum_i e^{-\frac{\beta}{2}\sum_{j_i}\left\langle h-t(\theta)\middle| K_{j_i} \middle| h-t(\theta)\right\rangle + \beta F}}$$

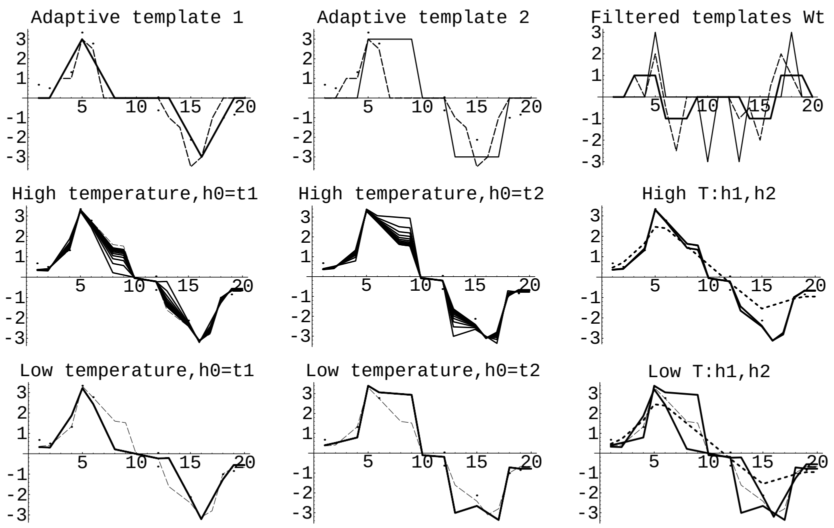
Now treat

☞ $\int d\theta$ by saddle point approximation

☞ $\sum_i$ exactly



Algorithm:
1. Find optimal θ^* in data space
2. Solve for h^* for fixed θ^*
3. Adapt θ^* for fixed h^*
4. Iterate till convergence.



Example: Optical flow fields

The stationarity equations for density estimation are nonlinear from the beginning due to a non–quadratic data term $\sum_i \ln P_i$ and the normalization condition.

$$\boxed{-\beta E_\phi = \langle \ln P[\phi] | N \rangle - \frac{1}{2} \left\langle \phi \middle| \mathbf{K} \middle| \phi \right\rangle - \langle P[\phi] | \Lambda_X \rangle\,,}$$

$P[\phi]$ (nonlinear) mapping from ϕ to P

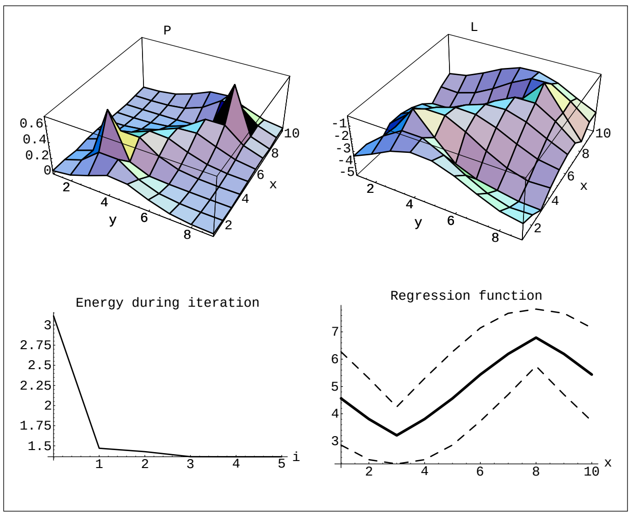
Λ_X Lagrange multiplier function

$$N(x,y) = \sum_i \delta(x-x_i)\delta(y-y_i)$$

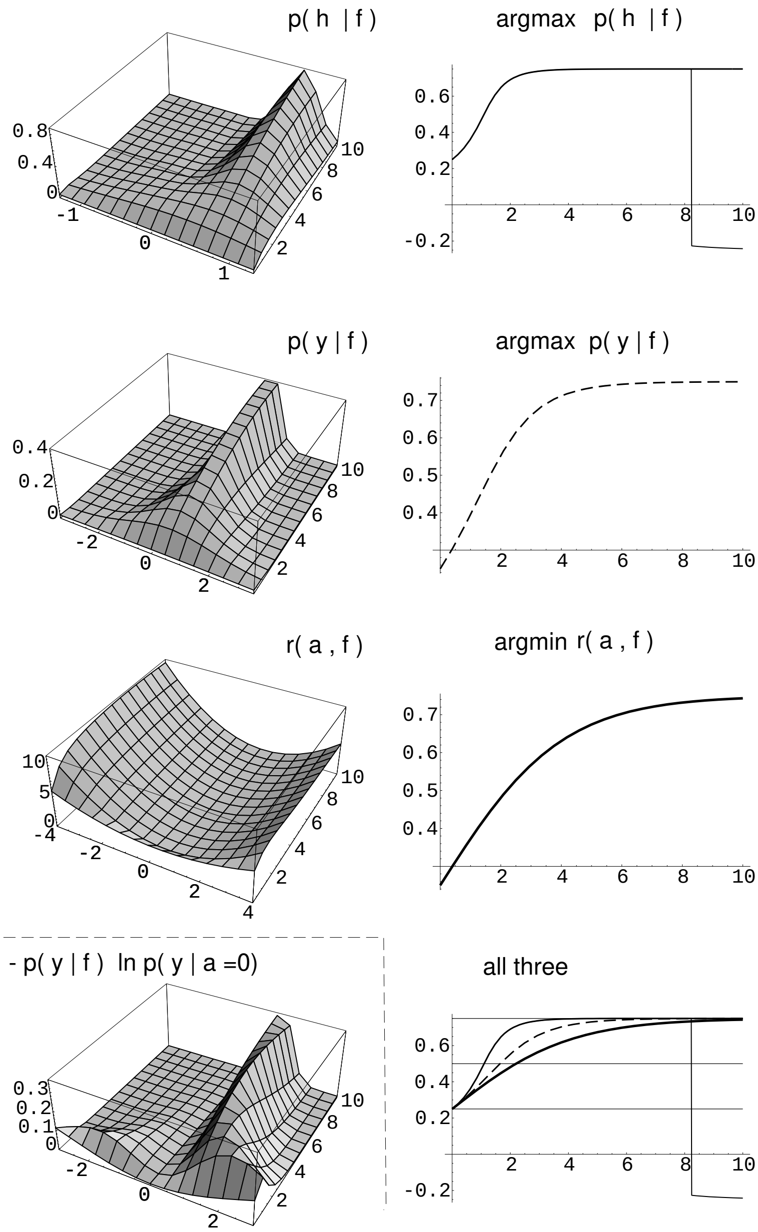
$$\boxed{\begin{aligned} &0 = \mathbf{P}'[\phi] \left(\mathbf{P}^{-1}[\phi] N - \Lambda_X \right) - \mathbf{K}\phi \\ &\Lambda_X = \mathbf{I}_X \left(N - \mathbf{P}\mathbf{P}'^{-1} \mathbf{K}\,\phi \right) \end{aligned}}$$

$$\mathbf{P}(x,y;x',y') = \delta(x-x')\delta(y-y')P[\phi](x,y)$$

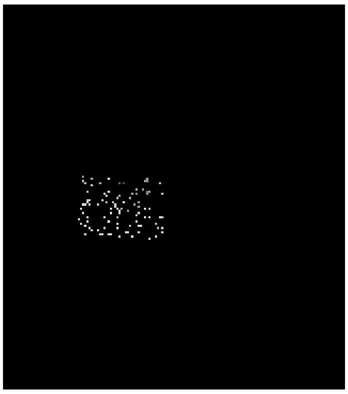
$$\mathbf{P}'[\phi](x,y;x',y') = \delta P[\phi](x',y')/\delta\phi(x,y)$$



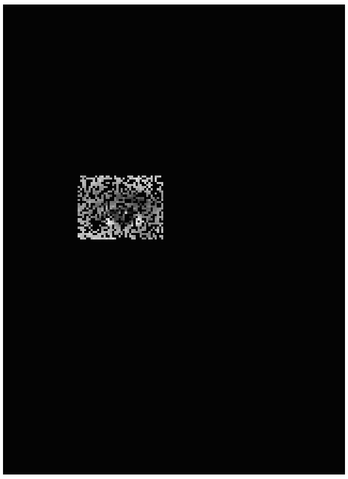
Maximum A Posteriori Approximation (Zero-dimensional example)



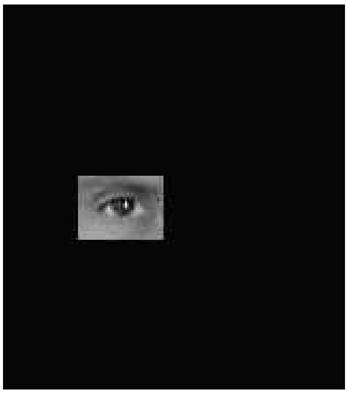
Some image data



100



1000



10000

Two adaptive templates

Adapted: Scaling, translation, rotation

Low temperature solutions

High temperature solutions

Yet another high temperature solution