Доказательство утверждения 2

Выполнил: Плетнев Никита

Если условное распределение не зависит от некоторого метапараметра, то добавление этого метапараметра к списку условий не изменит распределение и его характеристики. Поэтому можно переписать данные KL-дивергенции так:

$D_{KL}(q(w|\Gamma,\theta_1)||p(w,\Gamma|h_1)) = D_{KL}(q(w|\Gamma,\theta_1)||p(w,\Gamma|h_1,\lambda_1)) = D_{KL}(q(w|\Gamma,\theta_1)||p(w,\Gamma|h_1,\lambda'))$;

$D_{KL}(q(w|\Gamma,\theta_2)||p(w,\Gamma|h_2)) = D_{KL}(q(w|\Gamma,\theta_2)||p(w,\Gamma|h_2,\lambda_2)) = D_{KL}(q(w|\Gamma,\theta_2)||p(w,\Gamma|h_2,\lambda'))$.

Подставляя решения указанных задач, которые по условию являются единственными, получаем:

$\lambda^Q_{likelihood}E_{q(w,\Gamma|\theta_1)}\log p(y|X,w,\Gamma) - \lambda^Q_{prior\ 1}D_{KL}(q(w|\Gamma,\theta_1)||p(w,\Gamma|h_1,\lambda')) + \log p(h_1|\lambda_1) >$
$> \lambda^Q_{likelihood}E_{q(w,\Gamma|\theta_2)}\log p(y|X,w,\Gamma) - \lambda^Q_{prior\ 1}D_{KL}(q(w|\Gamma,\theta_2)||p(w,\Gamma|h_2,\lambda')) + \log p(h_2|\lambda_2)$ а также

$\lambda^Q_{likelihood}E_{q(w,\Gamma|\theta_2)}\log p(y|X,w,\Gamma) - \lambda^Q_{prior\ 2}D_{KL}(q(w|\Gamma,\theta_2)||p(w,\Gamma|h_2,\lambda')) + \log p(h_2|\lambda_2) >$
$> \lambda^Q_{likelihood}E_{q(w,\Gamma|\theta_1)}\log p(y|X,w,\Gamma) - \lambda^Q_{prior\ 2}D_{KL}(q(w|\Gamma,\theta_1)||p(w,\Gamma|h_1,\lambda')) + \log p(h_1|\lambda_1)$.

Сложим неравенства и отнимем одинаковые слагаемые от обеих частей, получим:

$-\lambda^Q_{prior\ 1}D_{KL}(q(w|\Gamma,\theta_1)||p(w,\Gamma|h_1,\lambda')) - \lambda^Q_{prior\ 2}D_{KL}(q(w|\Gamma,\theta_2)||p(w,\Gamma|h_2,\lambda')) >$
$> -\lambda^Q_{prior\ 1}D_{KL}(q(w|\Gamma,\theta_2)||p(w,\Gamma|h_2,\lambda')) - \lambda^Q_{prior\ 2}D_{KL}(q(w|\Gamma,\theta_1)||p(w,\Gamma|h_1,\lambda'))$.

Удалим лишние λ и перегруппируем слагаемые:

$(\lambda^Q_{prior\ 1} - \lambda^Q_{prior\ 2})D_{KL}(q(w|\Gamma,\theta_1)||p(w,\Gamma|h_1)) < (\lambda^Q_{prior\ 1} - \lambda^Q_{prior\ 2})D_{KL}(q(w|\Gamma,\theta_2)||p(w,\Gamma|h_2))$.

По условию множитель в скобках положителен. Делим на него и получаем:

$D_{KL}(q(w|\Gamma,\theta_1)||p(w,\Gamma|h_1)) < D_{KL}(q(w|\Gamma,\theta_2)||p(w,\Gamma|h_2))$, что и требовалось доказать.